ARI Research Note 95-43

Differential Assignment Theory Sourcebook

Cecil D. Johnson

Joseph Zeidner

George Washington University

for

Contracting Officer's Representative
Leonard A. White

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

Personnel and Training Systems Research Division
Zita M. Simutis, Director

July 1995

ARMY RESEARCH INSTITUTE

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

Edgar M. Johnson
Director

Research accomplished under contract
for the Department of the Army

George Washington University

Technical review by

Henry H. Busciglio
Peter Legree

Accesion For		
NTIS CRA&I		X
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 1995, July	3. REPORT TYPE AND DATES COVERED FINAL 9/91 - 8/93

4. TITLE AND SUBTITLE
Differential Assignment Theory Sourcebook

5. FUNDING NUMBERS
MDA903-91-C-0137
062785A
A791
2001
C05

6. AUTHOR(S)
Cecil D. Johnson and Joseph Zeidner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
George Washington University
Office of Sponsored Research
2121 I Street, NW, Suite 601
Washington, DC 20052

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-RS
5001 Eisenhower Ave.
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 95-43

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words):

Differential Assignment Theory (DAT) is presented as an alternative to other current theories that pertain to personnel selection and classification, but, unlike DAT, do not provide a basis of optimism for the successful development and implementation of both selection and classification-efficient operational systems. Data focuses on the research and development of systems that can effectively accomplish: (1) selection from a common pool of applicants, and (2) the subsequent optimal assignment of selected individuals to one of a number of alternative job families. The other theories at least implicitly assume that separate applicant pools exist for each assignment destination, thus permitting the evaluation of test batteries and assignment composites in terms of incremental predictive validity, essentially ignoring the effect of the intercorrelations among selection and assignment variables. DAT is described in terms of its assumptions, concepts, and the more than 30 principles that have been hypothesized and partially tested within the context of research on DAT relevant to selection and /or classification of personnel. The authors believe that true or more accurate descriptions of the interrelations among selected variables particularly relevant to selection and classification of personnel, including system, predictor, and criterion variables, are reflected in these principles. This report provides a source of such facts and concepts useful to the design of both research efforts and operational systems that have potential for the improvement of selection and/or classification policies, strategies, procedures, and total systems.

14. SUBJECT TERMS Personnel selection and classification ASVAB Model sampling experiments
Classification and assignment Army personnel policies Army operational aptitude areas
Differential assignment theory Army job families Mean predicted performance (Cont.)

15. NUMBER OF PAGES 76

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN Social1intelligence

StandardescriptionsRev 2-89)

Prescribed by ANSI Std 239particular

ARI Research Note 95-43

14. SUBJECT TERMS (Continued)

Predictive validity
Project A
SQT
Simulation of personnel selection and classification processes

FOREWORD

This report is one of a series of research efforts designed to improve the selection and classification efficiency of the Armed Services Vocational Aptitude Battery (ASVAB). The report defines differential assignment theory (DAT) in terms of its concepts, assumptions, and principles that are being developed on the basis of a number of empirical comparisons of DAT with alternative theories and approaches.

This report describes a large number of principles in six major categories: (1) models of classification efficiency; (2) clustering jobs into families; (3) selecting predictors for inclusion in an operational battery; (4) strategies; (5) alternative designs for operational selection/classification systems; and (6) criterion characteristics for classification.

DAT principles applicable to the construction of new and improved Aptitude Areas include: (1) the best test composites for either selection or classification are least squares estimate (LSE) composites; (2) an increase in assignment composite size provides a steady increase in classification efficiency as measured by mean predicted performance (*MPP*); and (3) an expansion of the number of job families to between 15 and 20 or more and clustering jobs into classification-efficient job families provides a substantial gain in classification efficiency.

The principles of DAT can now be applied to make changes in the Army's operational classification system in phases. The first phase of change may include: (1) the use of least squares composites (LSEs) for the existing aptitude areas as the assignment variables (AVs) for job families; (2) the use of either five-test composites or of tailored nine-test LSEs for each AV; and (3) the use of 15 to 20 or more classification-efficient job families and their corresponding AVs to substitute for the existing 9 job families.

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

The authors would like to thank the staff of the Selection and Classification Technical Area of ARI for their contributions to this research. The Contract Officer's Representative for this effort was Dr. Leonard A. White of ARI. The authors also would like to thank Dr. Peter Legree of ARI and Dr. Dale R. Palmer for their comments on the draft report.

DIFFERENTIAL ASSIGNMENT THEORY SOURCEBOOK

EXECUTIVE SUMMARY

Requirement:

Differential Assignment Theory (DAT) is presented as an alternative to other current theories that pertain to the design and implementation of both selection and classification-efficient operational systems. DAT is described in this report in terms of its assumptions, concepts, and principles that have been hypothesized and tested within the context of research on DAT. This report provides a source of concepts and facts useful in the design of research efforts and improvements in the total operational selection/classification system, including policies, strategies, and procedures.

Procedures:

The current state of knowledge concerning DAT principles is highly influenced by the theories of Brogden and of Horst published in the period between 1946-1964 and also by a series of model sampling experiments conducted at ARI for a decade starting in 1965 and then continued at George Washington University (GWU) since 1987 to the present. The theoretical framework and research results provide the basis of the current principles described in this Sourcebook. The findings of research now in progress should provide new data for an updated edition of the Sourcebook detailing additional DAT principles and knowledge.

The principles describe in the Sourcebook to date are based on studies that measure benefits in terms of mean predicted performance (*MPP*) and simulate the classification and assignment process using synthetic scores. Ongoing GWU research includes experiments that simulate S/C systems using either synthetic or empirical scores. The research methodology surveyed in this report compares a priori, *g*-based, and tailored test composites that generally allow free play or sampling error; or, in the case of a couple of studies, notes the effects of sampling error.

Findings:

The principles confirm the predictions of Zeidner and Johnson (1991b) that the use of efficient test selection procedures for a test battery or assignment composites along with the use of least squares estimates for test in assignment composites can greatly improve the utility of the Army classification process. A major increase in the number of assignment composites and their associated job families between 15 and 20 (and most likely more) would provide additional sizable gains in classification efficiency. Optimal classification can provide more than twice as much gain in predicted performance as gain obtained from selection alone.

Utilization of Findings:

The principles have a number of significant operational implications concerning the interrelations among systems, predictor and criterion variables, and concerning how samples and sets of predictor variables should be selected for analysis and used in design of composites. A review of the research studies that form the basis of the principles provide compelling evidence that there is a higher classification efficiency inherent in the ASVAB than is usually posited. The existing operational test composites could be redesigned to substantially improve classification efficiency. Included in redesign is the use of full nine-test least squares estimates for weighting of composites (providing maximum obtainable classification efficiency) or the use of five-test composites tailored to each job family, and an increase in the number of job families and in job family homogeneity.

In Zeidner and Johnson (1991b), we proposed a series of changes in each of nine areas. We indicated that the implementation of some changes can be made immediately, that some other changes require system development and testing before implementation, and that still others require additional research information to obtain estimates of gain and more precise specification of parameters. The implementation of an ideal operational classification system is unlikely to be accomplished in a single step. Traditions relating to classification systems and the administrative complexities involved in implementing changes inhibit making one overall change in the operational classification system incorporating all desired improvements. We therefore proposed a sequence of change over three time periods (pp. 197-211).

In early 1995, the technical details required to effect the first phase of an operational implementation will be available. This first-phase change should include better selected and weighted test composites for use as Army aptitude are composites associated with 15 to 20 or more job families.

DIFFERENTIAL ASSIGNMENT THEORY SOURCEBOOK

CONTENTS

Page

Introduction

A. Purpose of DAT Sourcebook

DAT is defined in this sourcebook by a set of basic assumptions and concepts. DAT is further characterized by principles that can be generated analytically from the basic assumptions and concepts of DAT or derived empirically. These DAT principles provide a basis for the design and development of effective operational selection and classification (S/C) systems, and for the conduct of research to confirm, modify, or reject these principles.

DAT was initially proposed by Johnson and Zeidner (1991a, 1991b) as a means of organizing the current state of knowledge about S/C systems into a more useful form: (1) by using mean predicted performance (*MPP*) as the figure of merit for the comparison of either alternative approaches or experimental conditions; and (2) by assuming that multiple job systems draw from a common applicant pool. The concept of DAT is derived from: (1) an integrative review of personnel selection and classification literature, especially the contributions of Brogden (1951, 1959, 1964), and Horst (1954, 1955); (2) analytic derivation of models and methodologies for improving classification efficiency by Johnson and Zeidner (1991); and (3) the findings of experiments involving the simulation of S/C systems and the output of *MPP* for the comparison of alternative conditions. DAT was immediately put to use as the conceptual basis for the design, conduct, and interpretation of research directed at the identification of features whose implementation would improve S/C systems.

Differential assignment theory is based on concepts of classification efficiency and differential validity first introduced by Brogden (1946, 1951, 1954, 1955, 1959, 1964) and Horst (1954, 1955, 1956). Their work has been integrated with and further extended through analytical derivations to form the set of concepts and principles that constitutes DAT (Johnson & Zeidner, 1991; and Zeidner & Johnson, 1994). Predictive data generated on the basis of this theory can be contrasted with data predicted by general ability theory, specific aptitude theory, and validity generalization theory.

Relating variables of S/C systems to *MPP* in the context of a common applicant pool for multiple jobs contrasts sharply with the practice of most current investigators who assume specific applicant pools for each job and use predictive validity as their primary figure of merit for evaluating batteries of tailored tests. The implicit assumption of independence of the applicant pools for each job in the sets of multiple jobs to which selection and classification is being addressed leads to different conclusions than DAT that insists that the classification efficiency (CE) of multiple job systems can be measured validly only in the context of a common applicant stream. The assumption of a common applicant stream greatly increases the probability that sets of tailored test composites will prove to be superior to *g* for either selection or classification to multiple jobs.

The value of test batteries cannot be assessed until their purpose has been defined. This handbook is restricted to the purpose of selecting and/or classifying personnel in a <u>common applicant stream</u> used for both selection and classification to multiple jobs after preliminary selection from the applicant pool. Selection to multiple jobs from <u>independent applicant pools</u> for each job is outside the scope of this sourcebook; this latter purpose has received nearly all of the research attention during the last decade, has been well covered in the research literature, and does not have the same dire need for attention at this time. We have been unable to locate a single research report on selection to multiple jobs from a common applicant pool that was published during the period between 1974 and 1989.

DAT predicts an increase in the mean predicted benefit of systematic selection and classification from the use of tailored test composites (LSEs) for selection and assignment. This prediction has specific implications for the way in which composites should be constructed. It also suggests that different approaches may be appropriate for the purposes of selection and classification. Building on the arguments of Brogden (1951, 1959) and Horst (1954, 1955), DAT argues for the use of tailored tests in operational test batteries which are selected to maximize differential validity. The theory further predicts a positive relationship between the number of tests in an operational battery and the mean predicted performance gain when all variables used to assign an applicant group are optimal least squares estimates. Hence, the larger the number of tests in a classification battery, the greater should be the gain in performance.

DAT's predictions contrast with the predictions of the adherents of a current mixture of *g* theory and validity generalization concepts, the more commonly accepted theory and practice in selection and classification. The *g* based theory has endorsed the use of assignment composites comprising tests selected to maximize predictive validity in a back sample and, as a direct consequence, emphasizes a single measure of general cognitive ability (g). Theorists who argue that the same measures are appropriate for selection and classification also usually regard the amount of incremental predictive validity over *g* provided by additional measures as the relevant basis for determining if anything other than cognitive ability is required for the construction of assignment composites. One result of this argument has been the acceptance by many investigators of aptitude composites designed to be primarily measures of *g*, based on tests measuring cognitive ability, and perhaps one or two measures of perceptual or psychomotor ability, as sufficient for classification (e.g., Hunter, 1984; Schmidt, Hunter, & Larson, 1988).

B. Definition of and Introduction to DAT

The conclusions reached from DAT-based research have thus far been presented in terms of the average least squares estimates of performance on the job for which an applicant has been selected and/or assigned. The mean predicted performance (*MPP*) of the selected and assigned group is scaled to have a mean of zero and a standard deviation of one in the applicant population. Such a least squares estimate of performance computed for selected personnel, using *MPP* as the figure of merit is called selection efficiency. When computed after both initial selection and subsequent optimal assignment to multiple jobs, the *MPP* index is called utilization efficiency. Utilization efficiency minus the effects of selection is called classification efficiency. Whether obtained by selection, classification, or by both, a given value of *MPP* provides the same economic benefit and can be used as one component of a measure of utility in comparing strategies that involve differing mixes of selection and classification.

The model sampling research paradigm is the heart of the DAT research method. Dat research typically focuses on the evaluation of selection and/or classification efficiency which may vary with differing features of a personnel system: test battery content, types of test composites, numbers of jobs or job families used in the assignment process, and selection versus classification strategies employed. Most DAT research results are obtained in the context of a two phase system in which selection into the organization is first accomplished using a single *g*-type composite and then assignment is made to specific jobs using weighted test composites tailored for each job.

Brogden and Taylor (1950) introduced the concept and usefulness of *MPP* as the basis of a dollar criterion for determining the value of selection. Brogden (1959) also proposed the use of *MPP* as a measure of the classification efficiency of a test battery in making optimal assignments to m jobs. Brogden provides a model in which $MPP = f(m)\, R\, (1 - r)^{1/2}$, where R is the mean predictive validity of the tailored tests and r is the mean intercorrelation among the tailored tests. These test composites are stipulated to be least squares estimates of job performance. Brogden provided a limited table for the order statistic, $f(m)$, that shows the gain obtainable, in standard score form, from assigning each person to the job corresponding to his/her highest criterion score.

The simulation based method for estimating *MPP* in the context of DAT makes considerably fewer assumptions than does Brogden's model, although conceptually *MPP* remains the same and has a mean of zero and a standard deviation of one in the applicant population. The use of least squares estimates of performance for each job as the surrogate for a direct measure of performance is vital to both Brogden's and our conceptualization of classification efficiency.

It is rare that performance scores for each of multiple jobs are available for each individual to be assigned. Brogden (1955) proved in an analysis of the classification process and computation of *MPP*, where predicted performance scores are defined as the least squares estimates of performance based on all predictors in a battery, that the predicted performance scores for each job could be substituted for the actual performance scores without changing the expected value of the obtained *MPP*. Abbe (1968) conducted a model sampling experiment showing that Brogden's proof of this important theorem was highly robust with respect to his primary assumption.

Horst's (1954) index of differential validity also substituted predicted performance for actual performance for multiple jobs. Johnson and Zeidner (1991, pp.106-107) have shown that Horst's differential validity index, H_d, is proportional to Brogden's *MPP* index when all of Brogden's assumptions are met and the number of jobs is held constant. Brogden and Horst's indices are primarily useful for selecting tests for inclusion in classification batteries and for obtaining preliminary estimates of classification efficiency in planning phases of system design. However, neither index is proposed by us for use as an evaluation measure of classification efficiency in the experimental comparison of systems or strategies.

Johnson and Zeidner (1991) show that H_d is biased by the presence of differences in validities and/or job values across jobs and cannot measure classification efficiency accurately under certain conditions. Brogden's *MPP* index has very stringent assumptions that are never met in real situations. Thus *MPP* obtained from a simulation is recommended as the figure of merit (or objective function) for comparing alternative policies, strategies, or systems.

C. Need for DAT

A misreading of the assumptions and basic concepts of those who propose the use of tailored test composites in multi-job S/C systems has permitted a number of *g* theorists and validity generalization proponents to depict the use of tailored tests as a very weak and vulnerable alternative to the total reliance on *g* in a selection process (Ree and Earles, 1994). To this end, an exaggerated requirement for independence among tailored composites, as well as from *g*, has often been described as essential to effective classification systems. The use of incremental predictive validity as the figure of merit with which to compare the contribution of tailored and *g* based test composites is another means used by both *g* theorists and many validity generalization proponents to discredit use of tailored test composites.

Brogden has erroneously been credited with (or accused of) being the principal proponent of a differential aptitude theory based on the supposition that tailored test composites are non-trivially more valid than test composites measuring *g* (Welsh, Kucinkas, and Curran, 1990). These authors equate differential aptitude theory to differential classification theory, describing both in terms of incremental predictive validity. They attribute the theory to Brogden and further compare it to the discredited theory of situational specificity, stating that Brogden assumed that "specific abilities can be measured and assessed for prediction of situational specific criteria." (p. 20) Of course, neither Brogden nor other DAT proponents believe that tailored test composites could be justified on the basis of incremental predictive validity, or that situational specific test composites were either desirable or practical.

Welsh, Kucinkas, and Curran (1990) cite differential classification theory as the basis of Brogden's contributions to personnel classification theory and practice and define this theory:

> According to the theory of differential classification, if each aptitude composite's validity is maximized in terms of its absolute validity, then there will be a maximization of the predicted performance of individuals within a cluster of specialties using the given composite. *The maximized predicted performance of jobs will in turn lead to maximized differences between job clusters in predicted performance, thus maximizing the differences in validities between clusters of jobs with differing composites* (differential validity).... It [theory of differential classification] assumes that specific abilities can be measured and assessed for prediction of situational specific criteria. (p. 20) [italics added]

In his many contributions to the scientific literature of personnel psychology, Brogden avoided using predictive validity, R, in the personnel classification context, other than as one term in his *MPP* function. He certainly never used R in the manner described above in italics.

A similar definition of differential classification or assignment theory erroneously attributed to Brogden is also provided by Schmidt, Hunter, and Larson (1988):

> Differential aptitude theory (or specific aptitude theory) postulates that specific aptitude factors assessed by particular tests or clusters of tests *make an incremental contribution to the prediction of performance* over and above the contribution of general cognitive ability. (pp. 1-2) [italics added]

Schmidt et al. (1988) further contend, citing Hunter (1983, 1984, 1985), that:

> ... based on very large military samples [research] appears to indicate that general cognitive ability is as good or better a predictor of performance in training in most military job families as ability composites derived specifically to predict success in particular job families. *These findings are contrary to the current theory that is the foundation of differential assignment of personnel to jobs in the military.* That theory, differential aptitude theory (or specific aptitude theory), postulates that specific aptitude factors assessed by particular tests or by clusters of tests make an incremental contribution to the prediction of performance over and above the contribution of general cognitive ability. (pp. 1-2) [italics added]

Hunter challenges the usefulness of job specific composites as resulting in, "composites that differ only trivially from the composites that best estimates general cognitive ability" (p. 356).

Hunter (1986) further challenges specific aptitude theory, stating that:

> A massive data base gathered by the U.S. Employment Service and even more data gathered by the U.S. military have shown the specific aptitude hypothesis to be false. (p. 358)

While it is clear that specific aptitude theory bears on simple selection for a single job, or selection for multiple jobs with independent applicant streams, we claim that a theory dependent on incremental predictive validity to evaluate tailored tests has virtually no relevance, other than to be misleading, to either selection from a common applicant stream for multiple jobs or for personnel classification. Specific aptitude theory, like situational specific theory, has been defined mainly by its detractors. It is not surprising then that a theory defined by those planning to discredit the hypotheses of this theory is both psychologically and psychometrically faulty and easily discredited when compared to general ability theory.

Incremental (predictive) validity is an important, often used, index bearing on the utility of adding an additional measure to a baseline variable to form an augmented variable. This index is defined as the increase in predictive validity provided by the use of the augmented variable(s) in addition to the baseline variable. In an associated research paradigm, one we refer to as the incremental predictive validity paradigm, the augmented variable(s) is (are) hypothesized to provided no increase in predictive validity over that of the baseline variable. When the research results do not lead to the rejection of this hypothesis, the alternative hypothesis is accepted and it is argued that the augmented variable does not provide greater utility than the baseline variable by itself. There is no reason to object to the use of either this index or the associated research paradigm when selection is for only one job or selection for each of multiple jobs is from independent applicant pools.

There is very heavy use of the incremental predictive validity paradigm in the current selection and classification research literature; the baseline variable is usually g and the research objective is frequently to show the adequacy of g for predicting the criterion without the help of additional variables, thus ostensibly supporting the sufficiency of unidimensionality as an explanation of empirically obtained validities.

This index and paradigm can be appropriate when selection for multiple jobs is the objective, even when the personnel being selected for each of the multiple jobs are drawn from a common applicant pool, when the hierarchical classification model is envisaged in the assignment algorithm (Johnson and Zeidner, 1991). However, since a single predictor composite is used in hierarchical classification to assign personnel drawn from a single applicant pool to multiple jobs using only a single assignment variable, an increase in incremental predictive validity depends upon the presence of varying levels of validities for the single predictive assignment composite for at least some of the jobs to which assignment is being accomplished. These varying levels may be due to differential validity of the composite across criterion variables but is more likely to be due to differing degrees of criterion predictability. Thus, when the incremental predictive validity paradigm is applied to the hierarchical classification model, there will be no gain in MPP due to the hierarchical classification process--over that obtainable from selection and subsequent random assignment to multiple jobs--whenever the single selection/assignment composite is completely lacking in the required variability in validity.

A more complex research situation presents itself when the objective is to determine the potential effect of adding additional assignment measures and corresponding assignment categories to the baseline measure, and effectively using the augmented set of assignment measures to select and assign to several additional assignment categories. In particular, we wish to consider a comparison of the baseline measure to the use of a set of tailored assignment composites to effect selection (and assignment) for the multiple job which corresponds to each tailored composite. Since the selection efficiency of each set of tailored composites depends on both the magnitude of their intercorrelations and their differential validities, in addition to their predictive validities, it should be obvious that the incremental predictive validity paradigm (which depends only on predictive validities) should not be used to compare the selection efficiency of *g* and a set of tailored composites. Yet, most of the literature on selection research relating to multiple jobs relies entirely on the incremental predictive validity paradigm to evaluate the potential effectiveness of sets of tailored composites.

The use of such a set of tailored composites, each corresponding to one job or job family, will in a simple selection process for multiple jobs have an impact on the effective selection ratio of all other jobs--unless each job is filled from an independent recruitment source. This assumption is essentially met for the following example of two jobs being recruited and filled at the same time: (1) one requiring technical courses and a college degree and, (2) the other requiring only a low level of reading and simple arithmetic skills. This independence is not a credible assumption with respect to the Army's recruitment of new enlisted soldiers.

While it sometimes appears that researchers who favor using the incremental predictive validity paradigm believe all differential validity across the tailored composites is due to random error, the substitution of the average validity across all jobs is not often relied upon. More often only the validities for the designated assignment composites are averaged. Some users of the incremental predictive validity paradigm may believe their use of the appropriate tailored tests or composites for obtaining the incremental validities has provided a measure of all the potential benefits that a set of tailored composites might be expected to provide. This confidence in the incremental predictive validity paradigm appears to ignore the greater increase in *MPP* that can be expected to result from decreasing the inter-correlation coefficient (r) as compared to increasing predictive validity (R) by one or two points.

To meet the need for an index of classification efficiency not embedded within the restraints of *g*-based theory, Johnson and Zeidner (1991) focused on the use of *MPP* in formulating a theory to serve as the basis for evaluating alternative selection and classification strategies. This theory, differential assignment theory (DAT), was also proposed in the hope of replacing specific aptitude theory as the basis for (allegedly) representing the views of

Brogden, Horst and other more recent proponents of the use of tailored tests in the multiple job situation.

Specific aptitude theory, as defined by validity generalization proponents and general ability theorists, could not possibly provide a basis for further research on personnel classification. Furthermore, apart from reference to any theory, a number of general ability theorists drastically misrepresent the views of those who espouse the use of tailored tests for classification. Thus the vital need for a theory capable of providing a strong alternative to the current marriage of *g* and validity generalization concepts is indicated. DAT provides this capability.

D. Origin and Development of DAT

The heritage of DAT includes the evolution of *g* based concepts of the structure of the intellect provided by a succession of theorists from Spearman's (1904,1927) general ability mental factor to Jensen's (1991) review of the predictive power of general ability. The multidimensionality of test space, so important to DAT, was popularized by Thurstone with his introduction of primary factors (1938,1947). More important, the factor analytic methodology also crucial to DAT was developed by researchers seeking the golden fleece of factor-pure tests. DAT applies the concept of multidimensionality of aptitudes in a joint predictor-criterion (JP-C) space to a concept of validity generalization of measures of predicted performance across job families.

With the first wave of popularity for personnel classification (Brogden, 1946, 1954), the idea of creating factor-pure tests for use in counseling and personnel classification appeared to some to be an attractive solution to the difficult task of developing new predictors that could provide meaningful classification efficiency. Many of those who shared in the mistaken euphoria of this earlier era over perceiving classification efficiency to be a relatively easy goal to achieve, may have been counting on the realization of "primary" factor-pure tests; proponents of such an approach assumed, much too optimistically, that correlations between factor pure tests would almost always be quite low in independent (cross) samples. Others may have assumed that the commonly believed requirement of low intercorrelations among assignment variables used in a classification process could be readily achieved.

Brogden's 1959 model shows that the correlations among assignment variables (AVs) used to make optimal assignments to jobs--the LSEs of predicted performance--can be as high as .95 and still provide a non-trivial amount of classification efficiency. More recently, research by Statman (1992, 1993) has shown that LSEs of factors rotated to simple structure in the JP-C space can be effectively used as assignment variables (AVs). When examined in independent (cross) samples such factor scores are less valid than LSEs of the job criteria but have sufficiently low intercorrelations essentially to compensate for the lower validities and provide equivalent mean predicted performance (*MPP*) in the classification situation.

In the decade from 1964 to 1974 the "SIMPO" research team of BESRL, now ARI, developed research techniques for evaluating alternative selection and assignment policies and procedures in terms of *MPP* computed at the end of a simulation of assignment to jobs. A series of model sampling experiments provided results that led this team to conclude that the computerized optimal assignment procedures being applied to a population largely of draftees was losing much of its potential gains in *MPP* because of the use of a number of ineffective strategies. These strategies appeared to stem from a lack of confidence among operational decision makers that utility could be obtained from personnel classification.

These ineffective strategies included, but were not limited to, the following: (1) conversion of Army Aptitude Area (AA) scores into the rough equivalent of an Air Force stanine score (instead of using as close to continuous scores as the raw test scores would permit); (2) definition of AAs in terms of 2 or 3 integrally weighted composites (instead of FLS composites); and (3) acquiescence to the inter-service trend towards eliminating of job interest and/or information tests from the classification battery and replacing these more differentially valid measures with tests saturated with *g* (Harris, 1966; Sorenson, 1965; Olson, Sorenson, Hayman, Witt, & Abbe, 1969; Sorenson, 1965). A number of technical reports supportive of DAT remain from the research efforts of the 1960's and 1970's, although the findings apparently have had little effect on the design and development of later S/C systems..

A period of approximately twenty years followed during which the work of validity generalization proponents and other *g* theorists led to a deemphasis of personnel classification. It is too soon to say whether a new era has now been initiated, an era wherein DAT fosters both optimism and interest in the potential utility of selection and classification systems.

The first phase of DAT development was the initial formulation of DAT described in the publication of four IDA reports (Johnson & Zeidner, 1991; Zeidner, 1987; and Zeidner & Johnson, 1991a and 1991b). These reports were based on the: (1) contributions of Brogden and Horst; (2) contributions of the BESRL-SRAD research team of the era 1964 to 1974; (3) the simulation studies of Nord and Schmitz (1991); and (4) several analytical studies described in Johnson and Zeidner (1991).

The second phase of DAT development was the completion of four experiments for which the need and tentative design was described in Zeidner and Johnson (1989, 1991b), and the results of a just completed experiment on selecting tests for job families described in this technical report. The results of the four experiments are reported by Johnson, Zeidner, and Leaman (1992, 1993); Johnson, Zeidner, and Scholarios (1990); Statman (1992); and Whetzel (1991).

The third phase of DAT development is ongoing research with ARI sponsorship that includes two experiments in a study initiated in 1991, two studies initiated in 1993, and a basic research study initiated in 1994--all being conducted by Zeidner, Johnson, and associates. The latter studies would: (1) explore multiple job selection and weight stabilization techniques; and, (2) develop improved LSE type test composites corresponding to job families. Research projects bearing on classification efficiency as conducted by AIR and HumRRO (and possibly others) in 1993, utilize much of the DAT technology and concepts. One of these efforts may, in the future, be viewed as the main stream leading into the fourth phase of DAT research.

The fourth phase of DAT development is longer-range research that includes: using more non-Project A populations; applying DAT to military vocational and educational counseling using an *MPP* output of systems simulations as the figure of merit; developing criterion composites that are more suitable for use in conjunction with S/C systems; incorporating MDS into S/C systems; drawing upon cognitive science to improve DAT and S/C systems; and developing a whole new domain of DAT to support the design of classification efficient computer-based adaptive testing systems.

E. Selection and Classification Myths

A number of beliefs prevalent among the *g* theorists, particularly among those who are also proponents of validity generalization, are described and referred to as "myths" by Zeidner and Johnson (1994). These "myths" would, if true, completely discredit the feasibility of designing effective multi-job S/C systems. It is clear that addressing these myths in a scientific but practical manner requires the application of a theory that incorporates the true positions of Brogden, Horst, and more current investigators who use research paradigms consistent with DAT. An ideal theory for this purpose should be capable of effectively structuring research on variables in S/C systems and of guiding the design of such systems.

In the following text we present examples of five common erroneous beliefs or myths of several eminent psychometricians and other investigators that appear in published journal articles or in reports of government sponsored research. The position taken by the cited investigators would, if true, prevent the design of effective personnel classification systems. Furthermore, the cited position is consistent with the investigators' more general positions, as known to us, on the use of tailored test composites for the selection and/or classification of personnel in the context of multiple jobs. All quoted investigators are too technically sophisticated to confuse selection and classification processes, but most appear to see nothing amiss in the evaluation of a classification battery in terms of selection efficiency. The positions taken by the cited investigators on these issues are orthogonal to the conclusions reached by DAT investigators on these same issues and show that a major controversy exists that should be dealt with from both an applied and a theoretical point of view.

1. The Dimensionality of the Joint Predictor Criterion Space

A belief in the unidimensionality of the joint predictor-criterion space is frequently stated in terms of a prediction that no differences would be found in the population among "best weights" for tests making up regression based composites. It is often assumed by supporters of unidimensionality that any deviation of test weights from equality (e.g., 1.0) in tailored composites is due to sampling error.

DAT contends that predicted criterion variance in a multi-job situation is typically multidimensional and that a pure measure of g (a measure which is equally valid for all jobs) is irrelevant to the classification and assignment of personnel to multiple jobs, and; is inefficient for selecting from a common applicant pool, as compared to tailored test composites.

Jensen (1984) provides an unmistakable message of the relative efficiency of *g* compared to other measures: "For most jobs, *g* accounts for all of the significantly predicted variance; other testable ability factors, independent of *g*, add practically nothing to the predictive validity" (p. 101). In his continuing discussion, Jensen recognizes the importance of spatial visualization and psychomotor tests for the skilled trades and perceptual speed and accuracy factor for many clerical occupations.

In reference to job criteria, Jensen further argues, "Specificity variance is probably plentiful. It contributes to the rather moderate ceiling, between 0.5 and 0.7 for test validity. But the prospect of devising tests whose cognitive specificity variance matches the specificity of any particular job is unfeasible and perhaps impossible. The specific 'factors' in cognitive tests, left over after *g* and two or three large group factors are extracted are inconsequential contributors to test validity" (p. 106). Jensen (1985) cites Cronbach as supporting his position with respect to the inadequacy of the ASVAB for effectively measuring other dimensions than *g*. "Cronbach (1979) has questioned the use of the ASVAB in educational and vocational counseling, essentially because the rather uniformly high *g* loadings of all of the subtests leave too little non-*g* variance to obtain sufficiently reliable or predictively valid differential patterns of the subtest scores for individuals" (p. 216).

Schmidt, Hunter and Larson (1988) also appear to believe that only the *g* component in the ASVAB makes a non-trivial contribution to validity:

> Recent research by Hunter (1983, 1984, 1985) based on very large military samples appears to indicate that general cognitive ability is as good or better a

> predictor of performance in training in most military job families as ability composites derived specifically to predict success in particular job families...the model that fits the data best is one in which the only ability causing performance is general cognitive ability and in which aptitudes are themselves caused by general cognitive ability... This theory of the underlying processes causing performance predicts that for military job families, general cognitive ability would predict performance at least as well as regression-based composites of specific aptitude derived to predict performance in the particular job family. (pp. 1-2)

The authors again make a similar point citing path analysis:

> Hunter's path analytic studies were conducted using average validities across all jobs for which validity data were available; these studies led to the prediction that general cognitive ability should have higher validity than regression-based composites of specific aptitudes for every job. (p. 4)

Ree and Earles (1991) use a different basis for concluding that weighted composites drawn from the ASVAB are unlikely to provide reliable measures of anything other than *g*. The authors interpret Wilks (1938) as providing:

> ... a mathematical proof that the correlation of two linear composites of variables will tend toward 1.0 under commonly found conditions... *g* may be found... by unrotated principal components, unrotated principal factors, or any one of a large number of possible hierarchical factor analogies, but also (up to scale) by any other reasonable set of positive weights. (pp. 276-277)

The authors conclude that the Wilks theorem makes these results predictable and generalizable to all measures of human cognitive aptitude that display positive manifold. It is only a small extension of this logic to assume that composites, with weights tailored to a specific job family, and containing tests drawn from a set of tests displaying a positive manifold, would also be just another measure of *g*.

Referring to the situation in which predictors are all measures of the same underlying general factor, a situation which the authors appear to believe to be highly prevalent, Hunter, Crosson and Friedman (1985) conclude:

> Optimal prediction is achieved... when the individual predictors are weighted according to the degree to which they correlate with the general factor" (p. 37)..." What may be surprising to some is the finding that General Cognitive Ability is the best predictor for all jobs. For all military data sets considered, the

path models were basically the same. The relationship between specific aptitudes and performance is causally mediated by General Cognitive Ability. (p.143)

2. Basing Both Theory and Design of Test Batteries on the Incremental Validity Paradigm

The second erroneous belief is that the usefulness of tailored test composites can be determined by the amount of incremental predictor validity that is provided by each of the tailored composites over that provided by a measure of *g*. When this increment is small, it is concluded that the use of tailored composites is unjustified. DAT offers MPP as an alternate evaluation index, one that is influenced by the possibility that sets of test composites possess differential validity to differing degrees and vary with respect to their population average intercorrelations among the composites.

We believe that those who hold this erroneous belief must implicitly assume that each job is filled from an independent source so that filling one job does not deplete the applicant pool available for other jobs. It is only when the assumption of independent pools can be justified that the use of incremental validity as an index for evaluation of the merit of tailored tests can possibly be appropriate.

This erroneous belief appears to be reflected in Hunter, Crosson, Friedman's (1985) statement that "the role of the Technical Aptitude Factor in the prediction of performance in all jobs is rooted in its *important incremental contributions* to the measurement of the General Cognitive Ability composite" (p. 92, italics added). The evaluation of the contribution of the technical aptitude factors in terms of its contribution to incremental validity is further discussed as follows: "Hunter's (1985) reanalysis of the Thorndike data suggested a fourth aptitude, Perceptual Aptitude, when added to the general Cognitive Ability composite increased the average validity from .59 to .61, an increase of about 3 percent. *Since the increase in work productivity is proportional to validity*, this would mean an increase of 3 percent gain in productivity" (p. 145, italics added).

This small increase in average validity is alleged to be an adequate measure of the improvement in all jobs obtained from adding another measure. This paradigm provides a valid insight into this improvement for the limited situation in which the personnel for each job are drawn from independent samples.

We believe average predictive validity is an inadequate means of assessing the value of an additional aptitude measure in a selection system where jobs are filled from a common applicant pool. The potential contribution of the additional aptitude requires determination of how much improved *MPP* can be obtained by using tailored test composites. Such an improvement is a function of the intercorrelations and differential validities of the composites, as well as of their predictive validities.

In summary, DAT prohibits the use of average predictive validity as the measure of the potential value of: test batteries, sets of test composites for use as assignment variables, new experimental predictors, or strategies for personnel selection, classification and assignment. Instead, the potential contribution of alternative selection/assignment system features should be determined from an index, such as *MPP*, which is consistent with use of a common applicant pool for multiple jobs, land the theoretical effect of both the interaction and the amount of differential validity of AVs.

3. Reliability of a Difference Between Composites

Hunter (1986) noting the high correlations among the ASVAB composites, states that:

> The only way to keep these correlations in the .80's or .90's is to restrict the number of tests in each composite and to artificially make the composites as close to non-overlapping as possible. Confirmatory factor analysis shows that these 'reduced' correlations are only artifactually lower than .95 because of error of measurement. If the correlations were corrected for attenuation, only the clerical composites would differ from the others.... A meta-analysis across hundreds of studies shows that the speeded tests make no contributions to the prediction of success in any occupational area except clerical and even there the contribution is minor (Hunter, 1985). (p. 356)

The effectiveness of tailored test composites (for the making of choices between a pair of jobs to identify the one with the highest predicted performance for a given recruit) is clearly dependent on the existence of a substantial reliability among the differences between the composite score pairs. Even when the intercorrelations exceed the reliabilities of both composites, this reliability of a difference has the potential of being non-trivially positive and can be argued to be usefully high. This potential is described further in the context of the description of a DAT principle.

The high correlation coefficients generally found among tailored test composites have often been cited as evidence that such test composites are not reliably distinct measures. Particular concern may be shown when the intercorrelations among composites exceed their reliabilities. The critics of tailored test composites sometimes cite the formula that Gulliksen (1987) provides for computing the reliability of a difference between two composites. This formula often indicates a zero, or even negative, reliability for differences among test composites. However, this particular reliability formula is inappropriate for use when the constituent tests of the composites on which the differences are based are not independent estimates across each pair of composites.

When the pairs of test composites contain overlapping tests, the reliability of the difference can be appropriately computed using a formula for correlation of sums between the composite difference and the true score difference. This formula yields respectable reliabilities for the differences among Army Aptitude Area composite scores (Zeidner & Johnson, 1994, pp. 391-392).

Hunter, Crosson and Friedman (1985) also illustrate both the first and third erroneous beliefs as follows: " When General Cognitive Ability is held constant, specific aptitudes do not add to the prediction of job performance... This poses problems for work in classification...this problem has been recognized for years in another form: *extremely high correlation between composites*" (p. 96, italics added).

Although it is always dangerous to guess why some one else comes to certain conclusions that are radically different than what you have reached, especially when it would appear that both parties have access to essentially the same empirical data, it is possible to make educated guesses as to why Hunter, and the many others who reach similar conclusions, conclude that the classification process has very little potential usefulness. We can summarize the possible reasons for their doubts in terms of their focus on the high intercorrelations among tailored test composites, the apparent instability of regression weight across independent samples, and their willingness to rely on the incremental predictive validity paradigm for the evaluation of the effectiveness of classification batteries.

DAT recognizes that intercorrelations among test composites in the .80 to .95 range can support a useful classification process (Brogden, 1959). Also, samples as small as 400 can provide effective regression weights for an assignment variable, although effective test selection requires larger analysis samples (Johnson, Zeidner, & Scholarios, 1994; more detail is provided by Johnson, Zeidner & Scholarios in press).

4. Standard Error of Regression Weights

A number of investigators appear to believe that test composites utilizing beta weights computed on moderately large analysis samples (e.g., N = 300) are patently useless for operational use. For example Hunter, Crosson and Friedman (1985) claim: "Since different aptitudes are highly correlated, very large samples are required for multiple regression. For good estimates of the population beta weights, samples of 5,000 or more are needed.... Consequently, the estimated beta weights tend to be far from their true values" (p. 18). Schmidt, Hunter and Larson (1988) agree with these conclusions regarding the instability of regression weights with the following statement. "Since intercorrelations among tests measuring the same aptitude would be high, computed beta weights would have large standard errors, and would be unstable from sample to sample.... if differential aptitude theory were valid at the level of general aptitudes but not valid at the level of specific aptitudes, *then beta weights for individual tests would be both unstable and difficult to interpret*" (p. 6, italics added).

Speaking of the ASVAB, Hunter (1986) says: "Ironically multiple regression on large samples leads to composites that differ only trivially from the composite that best estimates general cognitive ability [for an early statement of this fact see Humphreys (1962, 1979; for a recent meta-analysis, see Thorndike (1985)]. Meta-analysis has shown that nearly all of *the increase in multiple correlation due to using tailored composites has been due to sampling error*" (pp. 356-357, italics added).

The best evidence that assignment variables with multiple regression weights do in fact appreciably differ from general cognitive ability lies in the large increase in *MPP* provided by tailored test composites, as compared to that provided by a measure of *g*, when a DAT research paradigm is utilized in a research study. This superiority of best weighted test composites over the use of a measure of *g* occurs even when moderately sized analysis samples are used to compute the regression weights used for assignment variables, and *MPP* is computed in independent cross samples using evaluation weights that are independent of both the analysis sample and the cross samples. These results could not be obtained if Hunter and his colleagues were correct on this issue.

5. Psychometric Methods Best for Selection Are Also Best for Classification

The importance of the distinction between the assertion that selection efficiency in a test battery does not have to be decreased in order to achieve classification efficiency and a second assertion that the maximizing of selection efficiency will also maximize classification efficiency should be emphasized. The first assertion is frequently made by supporters of DAT. We believe the same moderately sized battery that has been specially selected to maximize classification efficiency will also be highly effective when used to maximize selection efficiency. It is unlikely that more than one or two tests, in addition to 7 or 8 tests selected to maximize classification efficiency, would be required to maximize selection efficiency.

Brogden (1946, 1951, 1959) has proposed the use of *MPP*, when computed after selection into an organization or to a single job, as a measure of selection efficiency. When computed after optimal assignment to jobs (or job families) he proposes using *MPP* as a measure of classification efficiency. Welsh, Kucinkas and Curran (1990) in their review of the ASVAB literature falsely accuse Brogden of originating a theory of "differential classification" that asserts the validity, rather than the classification efficiency, of tailored composites will be maximized for their corresponding clusters of jobs. As indicated in an earlier quote also found in the cited article, the authors state that such a maximization of selection efficiency will maximize classification efficiency.

The second of the above two assertions by Welsh et al. should mean that a battery for which tests are selected to maximize selection efficiency would also, necessarily, maximize classification efficiency. Research results provided by Johnson, Zeidner, and Scholarios (1990) and Scholarios, Johnson and Zeidner (1994) provide strong evidence that the use of an index usually associated with selection efficiency (i.e. a measure of predictive validity for multiple jobs) provides a test battery with inferior classification efficiency--when compared to a battery specifically selected to maximize classification efficiency.

While we believe that predictor validity is not even adequate as an index of merit for batteries with respect to selection for multiple jobs from a common applicant pool, it is evident that predictive validity is never an adequate measure of classification efficiency. However, this erroneous belief that maximum selection efficiency will also maximize classification efficiency may explain why Welsh et al. are led to describe Brogden's theory of differential validity in terms of predictive validity.

Assumptions and Concepts of Differential Assignment Theory (DAT)

A. Assumptions

DAT has been described in Johnson and Zeidner (1991) and more recently in Zeidner and Johnson (1994), in terms of a number of basic concepts and principles germane to both theory and operational applications. DAT can be succinctly characterized in terms of: the acceptance of *MPP* as the preferred measure of both selection and classification efficiency, and the adoption of predicted performance as a criterion variable in the multiple job situation.

In the long run a true assumption is something which is essentially not subject to empirical proof, either because of a lack of data or because of the immaturity of the investigative state-of-the art. A basic assumption of a theory is one which cannot be tested within the structure of that theory but which is essential to the generation of theorems or principles. The credibility of the theory rests upon its basic assumptions. DAT has only one such basic assumption, the substitutability of predicted performance for performance measures in multi-job selection and classification models. The other "assumptions" discussed below are subject to proof by DAT-based research methods that are ultimately based on empirical data.

The substitutability of predicted performance for performance measures in computing *MPP* for evaluating personnel systems, methodology, policies, and strategies in the multi-job situation, was already being utilized to develop classification models and methodology by Horst and Brogden when the latter provided his proof justifying this substitution (Brogden, 1955). Brogden necessarily made even more basic (but highly credible) assumptions in his proof of this basic DAT assumption. Abbe (1968), using a model sampling approach, established the robustness of Brogden's proof.

The remaining "assumptions" discussed below are not individually essential to DAT but have been usually accepted by DAT investigators seriously undertaking research regarding the selection and/or classification of personnel drawn from a common applicant pool for assignment to multiple jobs. The applicability of DAT to this kind of research requires consideration of all of the following concepts:

1. Reasonable care in selecting experimental test batteries can assure the presence of a non-trivial multidimensionality in both the reliable predictor space and the joint predictor-criterion space (JP-C).

a. Although the first principal component of the covariances among the predicted performances for each job family can be expected to range from 70 to 90 percent of the total explained variance, variance explained by the smaller components will usually make the major contribution to classification efficiency.

b. Multidimensionality in the predictor space is necessary but not sufficient to assure classification efficiency; differential validity across jobs or job families must also be present, i.e., multidimensionality in the JP-C space is required.

2. Even under the best of conditions as when appropriate selection of tests from either an operational or an experimental test battery for inclusion in tailored test composites comprising the assignment variables is followed by appropriate weighting of the tests selected for inclusion in each composite, the failure to use all tests in the battery will provide reduced dimensionality

in the JP-C space, as compared to the use of full least squares (FLS) composites as AVs. Also, the analysis samples for each job family used for test selection and computation of weights must be moderately large.

3. At least one important criterion component can be found or devised which will provide multidimensionality in the JP-C space. The finding of even one such component can provide adequate evidence of a major amount of potential utility that is obtainable from the use of a personnel classification system.

4. A model of a multi-job selection classification system, which assumes a common applicant stream, provides an appropriate representation of input into the military services and many other large organizations; in a large organization a common applicant pool (stream) is more likely to be found--as compared to an independent applicant source (stream) for each job.

5. The relationships among S/C system variables determined from moderately large samples of entities (vectors of synthetic scores) generated from a designated population can be expected to hold in the real population. Vectors of predictor scores obtained from data banks, when available in sufficient numbers, can be used instead of vectors of synthetic scores.

6. DAT requires consideration of the possibility that separate approaches may be required in designing an S/C system--depending on whether the objective is to maximize selection or classification efficiency. While Brogden (1955) proved that full least squares (FLS) test composites containing all the tests in the battery are optimal (in the back sample) for making either selection or classification decisions, DAT does not assume that both selection and classification efficiency can be simultaneously maximized as the result of all types of design or research decisions. In most S/C systems, potential classification efficiency (PCE) can be maximized with little or no reduction in potential selection efficiency (PSE). Other S/C systems can be maximized simultaneously for both PCE and PSE. Still other systems require a choice between maximizing PCE or PSE.

7. No inherent contradiction exists between validity generalization (VG) and DAT. The initial concept of VG as promulgated by Mosier (1951) can be fully incorporated into DAT. Many, if not most, VG concepts compatible with the selection and/or classification of applicants

from a common pool to multiple jobs, as introduced by recent proponents of VG, will eventually be incorporated into DAT. These concepts common to VG and DAT include many bearing on the importance of *g* for achieving predictive validity across all jobs: the high degree of generalizability of tailored test composites to jobs within a job family; the high generalizability of *g* to all jobs, when measured in terms of predictive validity; and the considerable, but lesser degree, of generalizability of tailored test composites to all jobs. All tailored test composites in operational personnel batteries (as known to us) will contain an impressive level of Brogden *g*, assuring a moderate level of validity generalization that has no effect on *MPP* potentially obtainable from optimal assignment and that is not even useful in the kind of hierarchical classification models that Hunter, Schmidt, and a few other VG investigators refer to as "placement."

A number of other assumptions which have usually been made by researchers while making use of DAT concepts or technology should not, in the opinion of the authors, be considered to be essential characteristics of DAT. These assumptions include the linearity of relationships between assignment variables and/or predicted performance measures and measures of utility, the equality of job values across jobs and job families, and the equality of utility differences across a given sized interval on the criterion scale at different levels. There is room for DAT based research on these issues.

B. DAT Concepts

Such DAT concepts as the non-trivial multidimensionality of aptitudes within the joint predictor-criterion space are subject to empirical verification. Since research hypotheses derived from DAT contrast with those derived from general aptitude theory, particularly with respect to the dimensionality of the joint predictor-criterion space, critical experiments bearing on the comparative credibility of these two theories can be readily devised. Thus, while unidimensionality has become an assumption for many theorists, dimensionality is a research issue, not an assumption, within DAT.

However, DAT's expectation of finding a non-trivial multidimensionality in the joint space is critical to the usefulness of personnel classification in operational systems. This expectation is challenged by some general aptitude theorists and validity generalization proponents.

Theories of the structure of intellect largely have focused on the factor structure of the predictor space while ignoring the importance of the joint predictor-criterion space. This has led to varying conceptualizations of general cognitive ability's role in predicting job performance--from the existence of a single general ability factor (Hunter, 1986; Jensen, 1986, 1991; Thorndike, 1985; Spearman, 1904, 1927) to the existence of multiple specific aptitudes which enable maximum overall predictive validity for jobs with different task demands (Fredericksen, 1968; Thurstone, 1935). Focus on the predictor space, however, is inadequate for explaining classification efficiency. DAT is entirely concerned with the multidimensionality of measures within the joint predictor-criterion space, as contrasted with either test space or common factor space.

The concepts and principles of DAT have special implications for the development and selection of tests for experimental and operational batteries. DAT, as described by Johnson and Zeidner (1991) and Zeidner and Johnson (1994), is consistent with findings of *g*'s dominant role in predictive validity. DAT, however, avoids justification, or rejection, for that matter, of tailored test composites on the basis of incremental predictive validity alone. Nor does DAT conform to situational specificity theory (Ghiselli, 1966). DAT theory, in contrast, holds that predictive validity is never an adequate measure of the value of tailored test composites nor of the battery from which the composites are derived. Only when predictors are to be selected for an operational battery, one which is to be used purely for selection from independent applicant groups for each job, is predictive validity the appropriate standard (figure of merit) for judging the value of a test battery.

The convergence of validity generalization theory (VGT) and differential assignment theory, as described above, requires only the substitution of moderate to highly correlated group factors for *g*, and relaxation of the faith of many VGT proponents that the weights of tailored tests are too unstable to permit classification effectiveness in independent samples. The authors cannot envisage a comprehensive DAT that does not incorporate all but a few of the many contributions of the validity generalization movement.

Most of the concepts that characterize DAT pertain to research methodology rather than to content methodology. DAT is not in general a "content" theory, although the use of DAT research methodology has potential for enriching our knowledge of psychological measures. Five of the more important DAT methodological concepts are discussed below.

1. The message of DAT is one of optimism regarding the possibility of designing, developing, and implementing improved S/C systems--ones with potential for being far superior to existing operational systems. It is believed by the authors that the deliberate application of DAT in the development and implementation of new S/C systems is the most practical and certain way to accomplish this goal.

This message could be referred to as an assumption, except that the truth of research results based on DAT do not rely on the accuracy of this message. This message is a basis for forming research hypotheses, not an assumption which must be accepted in order to accept DAT research results, or whose disproof could negate research findings obtained in the context of DAT. However, there is no point in conducting research on a classification system unless this message is accepted.

The converse of this central message of DAT is that optimal assignment to job families would not be a practical feature to consider installing into a personnel system if the central DAT message is not credible. There are a number of characteristics which would preclude a serious effort to develop an effective personnel classification system if they cannot be eliminated as a serious consideration. The following are among these critical characteristics: (1) unidimensionality of the joint predictor-criterion space; (2) so much error in the weights of tailored test composites that the composites are indistinguishable from each other in the population, and; (3) sampling error so overwhelming in all MOS clustering approaches as to prevent the forming of stable job families. Pessimism regarding any of these three issues virtually halts further work on personnel classification. However, we believe DAT can provide a useful research and development context because we accept the optimistic central message of DAT.

2. DAT prefers the maximization of utility values over psychometric indices; it is assumed that utility models, where the object is to maximize the benefits obtainable for a fixed cost, provide the best basis for evaluating alternative policies and procedures being considered for operational implementation. *MPP* provides a credible measure of benefits and permits the comparison of costs and benefits obtainable from many different processes, including selection, placement, classification, training, and recruiting. Thus the obtaining of *MPP* is a useful first step in computing utility.

3. To obtain values of *MPP* representing each experimental condition in research bearing on selection and/or assignment to four or more jobs from a common applicant pool requires simulation of the key aspects of the S/C system under each experimental condition. Most DAT-related research requires the simulation of the assignment to jobs of entities defined in terms of vectors of test scores. This simulation requires the inclusion of an optimal assignment algorithm when potential classification efficiency is being determined, or, alternatively, when an estimate of operational classification efficiency is being sought, an appropriately simplified version of the operational assignment process.

4. DAT can make important contributions to the design of operational S/C systems. Computer technology and costs have reached the state where it is practical to implement virtually any decision process that can be shown to provide a useful gain in *MPP*, regardless of the complexity of the algorithms and/or data structures required to implement the process.

5. DAT places its primary emphasis on the joint predictor-criterion space--that is, on a subset of the reliable space of the predictors that is shared with the criterion space. Factor analytic methodology is highly useful in the conduct of research, design, and development of S/C systems, as well as in the conduct of research on DAT principles.

DAT Principles

A. Basic DAT Principles Related to Classification Efficiency

1. Brogden's *MPP* model (1959) can be written as $MPP = f(m)\ R\ (1\text{-}r)^{1/2}$. Mean predicted performance (MPP) is a function of the number of jobs (m) to which individuals are optimally assigned, the (average) validity (R) of the least squares estimates (LSEs) of the job criteria used to make assignments, and the intercorrelation (r) among the LSEs. (*MPP* is obviously not just a function of R.)

2. The assumptions of Brogden's 1959 model can be expressed in terms of Spearman's two-factor model.

a. The loadings of each job predicted performance variable on Brogden's g can be computed using the algorithm described by Johnson and Zeidner (1991). Commence by first factoring the covariances among the predicted performance variables (**C**) to obtain the m by n factor matrix, **F**, where m is the number of jobs and n is the number of predictors used to compute the predicted performance variables ($m > n$). The matrix $\mathbf{C}_d$, obtained as indicated, is then computed: $\mathbf{C}_d = (\mathbf{F} - \mathbf{HF})$, where **H** is an operating matrix with all elements equal to $1/m$, with the same numbers of rows and columns as **F**. Next, obtain a principal component solution of $\mathbf{C}_d$. The last column of this factor solution, $\mathbf{F}_d$, is Brogden's g. If using a canned PC solution, be certain that the diagonal values of the covariance matrices are utilized unaltered.

b. Brogden's g, as computed in the joint predictor-criterion space, can be readily extended to the predictor space using another algorithm described by Johnson and Zeidner (1991, pp. 97-98).

3. The g as utilized in the expression of the assumptions in Brogden's 1959 model in terms of Spearman's two-factor model is a special kind of g defined as having equal correlations with all LSEs in a designated set of assignment variables (AVs). It is easy to demonstrate that Brogden's approximation of MPP ($\text{MPP} = f(m)\ R\ (1\text{-}r)^{1/2}$), and thus of classification efficiency, is invariant under the addition or subtraction of Brogden's g from the predictor variables, when the assumptions of his model are met. (Brogden, 1959: Johnson & Zeidner, 1991) When we compute the effect on MPP, defined as above, of altering the amount of Brogden's g without using any particular factor theory to define the relationships among the model variables, we find

that increasing the amount of Brogden's g under these more relaxed assumptions lowers the estimated *MPP*.

a. The elimination of Brogden's g has no effect on the potential classification efficiency (PCE) as computed in a "back" sample when Brogden's g is removed from the assignment variables. See Appendix A for further explanation.

b. The removal of Brogden's g from the assignment variables (LSEs) can be expected to reduce the average standard error of the regression weights found in these LSEs, possibly increasing PCE in cross samples. See Appendix B for further explanation.

c. Brogden's g has a much smaller validity than does either the first principal component or psychometric g. Brogden's g can be readily computed by orthogonally rotating the principal component solution of the covariances among the LSEs to successively maximize the value of H_d for each rotated component. The last component will have equal correlation coefficients with all predicted performance measures (i.e.,LSEs) and is Brogden's g. This method for obtaining a solution to maximize H_d is described in Johnson and Zeidner (1991, pp. 101-104).

4. Horst's index of differential validity (H_d) is consistent with the measure of *MPP* in Brogden's 1959 model. When Horst's index of differential validity (H_d) is computed from a two-factor solution, H_d is proportional to *MPP* as defined by Brogden's model when all of Brogden's assumptions are met. If the number of jobs (m) is held constant *and all of Brogden's assumptions are met*, the rank order of the output resulting from a number of different experimental conditions (e.g., alternative sets of tests or jobs) will be the same whether the output is computed using H_d or *MPP* as the figure of merit.

5. A principal component (PC) solution of the covariances among the LSEs (predicted performance variables) used as assignment variables successively maximizes Horst's absolute validity index (H_a) for multiple jobs.

a. H_a is an index of predictive validity; H_a is the average squared multiple correlation coefficient in which all predictors are used to form LSEs for each job criterion.

b. The first component from the PC solution which maximizes H_a provides considerable less PCE than does the first component from the rotated solution which maximizes H_d.

6. The restriction in range effect of first stage selection augments the average PCE for the assigned individuals resulting from second stage classification.

a. Whetzel (1991, pp. 43-45, 90-92) shows that the *MPP* of optimally assigned individuals is higher for a selection ratio of .50 when simulataneous selection and assignment is utilized than for a selection ratio of .75, even when the amount of *MPP* due to selection is removed prior to making the comparison. It is easily demonstrated that this is the expected result in an analytical solution of this problem. See Appendix C for further explanation and Appendix D for required computational techniques.

b. It is readily seen that restriction in range effects from the use of a measure of *g* to effect selection of some applicants can be expected to have more secondary effect on the intercorrelations among tailored test composites than on the validities of these assignment variables. The reduction in the magnitude of *r* will often result in double the increase in magnitude of $(1-r)^{1/2}$ as compared to the reduction in *R*. The product of $R(1-r)^{1/2}$ can thus be expected to increase as the restriction in range effects from selection increase. The reader can readily demonstrate this to himself by assigning values to both *R* and *r* and then separately providing the same increment to one variable.

c. Within a highly selected group, such as Army special forces, further classification to MOS might well provide a greater gain in *MPP* than is obtainable in an assignment process taking place on the total Army input (i.e., the initial assignment process).

B. DAT Principles Related to Clustering Jobs Into Families

1. When allocation effects are present in the assignment process, the effect of the number of jobs or job families (*m*) on potential classification efficiency (PCE) does not level off as rapidly as predicted by Brogden's 1959 model; the optimal number of job families corresponds to the number of related job clusters for which validity data are adequate for computing stable LSEs for use as AVs (Scholarios, 1992; Leaman, 1992; Statman, 1992).

a. Brogden's 1959 model argues for a major role for the number of jobs or job families to which optimal assignment is being made (m) in the determination of expected MPP computed after optimal assignment to jobs. The increment predicted by this model is independent of changes in R and r occurring with an increase in the number of job families (as when existing job families are shredded to increase homogeneity within and heterogeneity across families).

b. Johnson, Zeidner and Scholarios (1990) found, under conditions where the number of jobs was varied in the absence of job clustering, that the increase in *MPP* resulting from an increase in number of jobs from 9 to 18 was less than predicted by Brogden's model; it was noted that the models' assumptions regarding an increment in dimensionality with the addition of each job to the assignment process was not being met.

c. Johnson, Zeidner and Leaman (1992) varied the size of m by either using a priori job clusters provided by the operational system, or by clustering jobs so as to minimize the reduction in H_d. Both methods increased homogeneity of jobs within families (increasing R) and heterogeneity of jobs across families within the JP-C space (decreasing r). A further effect can be assumed to have occurred as a result of the f(m) of Brogden's model. They found that *MPP* increased with an increase in m for as many job families as could be appropriately created with the available data.

2. The effect on *MPP* of increasing the number of job families by using an H_d-oriented algorithm for clustering jobs exceeds the increase in *MPP* provided by using operational families (CMFs) as the basis for increasing m; this algorithm commences by considering all jobs as separate job families and by reducing by one the total number of job families at each iteration, minimizing the reduction in H_d. The agglutination of jobs into a decreased number of job families continues until the desired number of job families is reached. (See Johnson, Zeidner, and Leaman, 1992.)

3. When predicted performance is used as an assignment variable based on a single variable (e.g., "g"), weighted by its validity (or value) for each job family, the number of job families has very little effect on CE. Such a classification model with total reliance on g is the

hierarchical classification (HC) model. The HC model is compared with the allocation model in Johnson and Zeidner (1991, pp. 29-41).

4. Consider clustering jobs by the following procedure: (1) first rotate factors to simple structure in the JP-C space; (2) assign jobs to a job family corresponding to the factor for which each job has its largest loading; (3) use a LSE of the appropriate factor as the assignment variable. This procedure provides a credible alternative to the use of FLS composites as assignment variables for job families identified in this manner (Statman, 1993).

5. Research findings support the practicality of designing two-tier classification systems, the first tier consisting of the maximum feasible number of job families with initial assignments determined by FLS composites, the second tier consisting of a smaller number of job families corresponding to factors for use in vocational counseling and setting minimum scores for acceptance into training or special programs (Zeidner & Johnson, 1991b, pp. 10, 205; Statman, 1993).

a. The use of factor scores as second tier AVs, where factors are rotated to simple structure in JP-C space, can closely approximate the CE obtained from the use of LSEs of criterion variables as first tier AVs (Statman, 1993).

b. We believe it would usually be possible to obtain the second tier job families without the necessity of splitting up job clusters obtained for the more numerous job families utilized in the first tier.

(1) First identify the second tier of job families; then identify the first tier families under the constraint that jobs together in one tier must be together in the other tier.

(2) First identify the first tier job families; then use these families, rather than the individual jobs, for the rotation to simple structure that permits identifying the second tier job families.

C. DAT Principles Related to Selection and Assignment Variables

1. Full least squares composites are optimal in the back sample for the accomplishment of either selection or classification, or both simultaneously.

2. The selection of predictors from an experimental pool *for inclusion in an operational classification battery* is best accomplished by the use of a measure of differential validity as the figure of merit in the test selection algorithm.

a. The best figure of merit reflecting differential validity is the point distance index (PDI), closely followed by the Horst differential validity index (H_d); both of these indices are superior to Horst's index of absolute validity (H_a).

b. PCE (i.e. *MPP* after optimal assignment) is significantly higher when a measure of *differential* validity (e.g., H_d or PDI) is used as the figure of merit in a test selection algorithm to form an operational test battery--instead of a measure of *predictive* validity (e.g, H_a or Max-PSE). Max-PSE is the Maximum Personnel Selection Efficiency index, and H_a is Horst's selection efficiency index--both for a multiple job (or job family) situation. Where H_a is essentially the average of the squared validity coefficients across jobs, Max-PSE is the average of these coefficients (unsquared).

c. It is easily shown analytically that Max-PSE is superior to H_a because the former is based on the average of multiple correlation coefficients with the job criteria in a multiple job situation, whereas H_a is similarly based on the average of the squared multiple correlation coefficients; H_d is also a squared concept while PDI is an average of non-squared functions that are squared and averaged to constitute H_d.

d. If the test battery is to be used only for selection, MPP is maximized by using H_a or Max-PSE as the figure of merit in the test selection algorithm. The selecting of tests for inclusion in a test composite has a fundamental difference in both objective and import as compared to the selecting of tests for inclusion in batteries (as in C2 above). The content of batteries should cover the dimensionality of all jobs (or job families), while the test composites, while limited to the content of the battery, need not cover criterion dimensionality other than that associated with a specific job or job family.

3. The separate selection of predictors from an experimental pool for inclusion in the test composites (when they are to be used as assignment variables to accomplish optimal classification) can be appropriately accomplished by a measure of predictive validity as the figure of merit in the test selection algorithm.

4. Most of the gain in CE obtainable from the use of assignment variables, which are best weighted test composites based on appropriately selected sets of tests, will remain when the tests are selected by a similar process that assures all the weights are positive (Johnson, Zeidner & Scholarios, in press).

a. The loss in CE resulting from constraining the test selection process to obtain a set of 9 tests from the 29-test battery yielding all positive least squares weights is approximately equal to one third of the loss resulting from decreasing the number of tests in each LSE type test composite from 9 to 5 tests and still permitting negative weights, when a 9 job family system reflecting the existing operational system is simulated.

b. The loss in CE resulting from constraining the test selection process to yield all positive least squares weights in obtaining a set of 5 tests from the 9-test ASVAB is approximately equal to one half of the loss resulting from decreasing the number of tests in each composite from 5 to 3 tests and permitting negative weights--when a 9 job family system reflecting the existing operational system is simulated.

c. The test selection constraint used to assure that all least squares weights in LSEs of job criteria are positive can also be applied to the selection of tests to be included in a set of tests to be used as LSEs of factors; thus, factor scores can also be defined in terms of least squares weights that are all positive. However, the reduction in *MPP* resulting from the use of only positive weights for LSEs of factors would be expected to be greater, as compared to the LSEs of criterion variables.

5. The gain in CE over chance assignment resulting from the use of k tests selected separately for each job family can be considerable when compared with the use of a single set

of k tests selected to constitute an operational battery. However, the number of tests that must be included in the operational battery also greatly increases when the experimental test pool is as large as 29, making it impractical to use this approach for directly selecting tests for composites (Johnson, Zeidner & Scholarios, in press).

a. The gain in CE obtained by separately selecting (using PV) the tests in each 5-test composite from a 29-test battery--as compared to the selecting of a single 5-test battery from which the AVs for each job family use all the tests with separate sets of weights--is half again as large (.074) as the gain (.051) obtainable by selecting (still using PV) and optimally utilizing a best 9-test battery. If the best 9-test battery is selected using H_d, the gain obtainable from selecting a 9-test battery is only slightly lower (.066).

b. The *gain* in CE from separately selecting the tests in each 3-test composite from a 9-test pool (the ASVAB)--over that obtainable from selecting (using PV) a single 3-test battery from this same pool--is .087. The AVs used in connection with each battery are FLS composites. The gain obtainable from selecting and optimally utilizing a 5-test battery is .073, and the gain obtainable by expanding the battery size to 9 is .110. The gain obtainable from both separately selecting each composite and increasing composite size to 5 is .101, compared to a gain of .096 obtained by using a 5-test battery (selected using H_d) and the use of FLS composites.

c. The gains from separately selecting tests for each AV requires a considerable increase in the size of the implied battery, as well as testing time, when selection is from the 29-test battery. When the experimental test pool contains 29 or more tests, the effect on CE of adding tests to the operational battery (selecting from a 29-test experimental test pool) does not level off until somewhere beyond 20 tests.

(1) Increasing battery size from 3 to 5 tests (tests having been selected from a 29-test Project A battery to maximize predictive validity), provides an average increase in *MPP* (per increment in the number of tests) of .0115 as compared to an average increase of .01275 when battery size is increased from 5 to 9; there is definitely no leveling off indicated in these findings.

(2) Comparing *MPP* obtained when battery size is 20 with the MPP obtained using the full set of 29 tests shows that the effect of battery size is still evident when n = 20 is compared with n = 29 using a research paradigm which controls all sources of correlated error.

6. The effect on CE of adding more tests to test composites from a 29-test pool does not appear to level off any faster than does the effect of adding more tests to an operational battery; available findings indicate that this effect does not level off until somewhere beyond 9 tests when selection is from a 29-test Project A pool (no evidence is available regarding the effect of larger composites on *MPP*).

a. Increasing test composite size from 3 to 5 tests (selecting tests from a 29-test Project A pool to maximize predictive validity), provides an average increase in *MPP* (per increment in the number of tests) of .013 as compared to an average increase of .012 when composite size is increased from 5 to 9; as in the above principle relating to the selection of batteries, there is no statistically significant leveling off indicated in the available data.

b. The selecting of tests from the ASVAB does show a leveling off of gain in *MPP* from adding tests to test composites. An average increase in *MPP* (per increment in the number of tests) of .007 is obtained as compared to an average increase of .002 when composite size is increased from 5 to 9.

7. Based on a model sampling study using the Project A concurrent validation study data, it appears that when PV test selection is utilized to select composites containing 3, 5, and 9 tests, selecting a 3-test composite from the 29-test pool provides no advantage over selecting from the ASVAB (gain = .001); for 5-test composites a very small advantage is provided from the use of the 29-test pool instead of the ASVAB (gain = .009). The only substantial gain from using the larger test pool as the source of the composites occurs when 9-test composites are being selected (gain = .042) (Johnson, Zeidner & Scholarios, in press).

a. When PV is the basis for selecting the tests in the LSE by a sequential accretion method, and the *selection is made from the ASVAB*, the best 5-test composite provides almost as much *MPP* as is provided by a FLS composite (the full ASVAB); the gain provided from using 9 instead of 5 test composites is .009 as compared to the gain of .014 provided by using 5 instead of 3 tests in each composite.

b. If all the tests in the ASVAB are used to form LSEs for use as AVs (i.e., new Army Aptitude Areas)--selecting tests in the same way as above, but with the constraint that only tests for which the least squares weights are positive are selected--the reduction in *MPP* is only .006; the reduction resulting from the application of this constraint to the selection of 5-test composites is .011. Thus the loss in *MPP* due to the selection of only positively weighted tests from the ASVAB for a 5-test composite is of a similar magnitude as in the reduction in composite size from 9 to 5 tests.

c. Again using Project A data and selecting tests in the same way as above, the same *MPP* is obtained for a 3-test composite whether tests are selected from the ASVAB or from the 29-test Project A experimental pool (.220 vs. .221); the gains obtained by adding tests to the composite are greater when test selection is from the larger pool, as compared to the ASVAB with a gain of .064 accruing from increasing composite size from 3 to 9.

d. When a k-test composite is compared with a k-test operational battery, it should be noted that a 5-test composite directly selected from the 29-test experimental battery will require an increase in size of the operational battery of at least twice that of the existing ASVAB. It is of considerable theoretical interest that the effect on *MPP* of adding to the number of tests directly selected for inclusion in each composite is in the same direction as the effect of adding more tests to a battery, although much smaller when the effect of increased battery size is considered.

8. Factor scores as AVs, when the factors have been rotated to simple structure in the JP-C space and job families have been identified in terms of these rotated factors, provide as much PCE as can be obtained by LSEs of the job criteria when computed within the same factor space and provide almost as much PCE as is obtainable from the use of FLS composites computed in the total joint test and criterion space.

9. Traditional weight stabilization methods applied to the weights in assignment variables (AVs) do not appear to increase classification efficiency (CE).

a. The literature on weight stabilization has focused entirely on the effect of methods that reduce the range of least squares weights on predictive validity. Most of the proposed methods appear intuitively to be counter productive with respect to classification efficiency.

b. The two most reliable methods for increasing stability of regression weights is to increase sample size and/or decrease the number of predictors. Four of the other traditional methods for weight stabilization are: (1) elimination of negative weights; (2) reduction in the number of predictors in a LSE through sequential test selection; (3) reduction in the dimensionality of test space by using only the larger of the principal components to define a reduced space within which to compute weights; and (4) use of unit weights.

c. The first three of the above four methods have been tried out in DAT-oriented simulation experiments with no instance of any of these types of AVs showing an increase in *MPP* over the use of FLS composites as AVs. The last method, use of unit weights, has provided significant reductions in *MPP* (Johnson, Zeidner, and Scholarios, in progress).

d. A large number of other methods proposed for weight stabilization and evaluated in terms of predictive validity in independent (cross) samples have not been tried out in a personnel classification situation.

10. The proportional size of the first principal component in the JP-C space, as compared to total factor contributions, is usually larger than the proportional size of the first principal component in test space. The present ASVAB can be described factorally as follows:

a. The first principal component in the JP-C space will typically provide 80 percent, or more, of total factor contributions.

b. The combination of principal components other than the first will almost certainly provide more classification efficiency than will the first principal component (Whetzel, 1992, Statman, 1992).

D. DAT Principles Related to Operational Strategies and Alternative Designs for Operational Selection, Classification, and Placement Systems

1. Hierarchical layering provides an analytical model for classifying personnel in accordance with predicted performance (PP) when PP is defined as an individual's score on a single measure (e.g., *g*) weighted by the measure of validity or value for each job or job family; personnel classification accomplished by means of hierarchical layering is referred to as hierarchical classification (HC).

a. Hierarchical layering with respect to validity consists of a process in which the job for which the AV has the greatest validity (or value) is filled with the selected applicants having the highest PP scores; the job with the second highest validity is then filled with the remaining applicants having the highest PP scores, etc., until all jobs are filled.

b. The hierarchical layering model will make the same assignments (or equivalent assignments where there are ties) to jobs as would be accomplished by using a linear programming (LP) model in which the objective function is PP.

c. The hierarchial layering concept, in which PP is defined in terms of tailored test composites, instead of in terms of a single measure weighted by validity or value, requires the use of an LP algorithm (or an equivalently complex optimization algorithm) to implement.

d. When tailored test composites are used in a LP algorithm to maximize CE, the effects of hierarchical layering (i.e. HC effects) and allocation are not separable--unless the tailored test composites have been standardized so as to eliminate all HC effects.

2. *MPP* can be increased in a personnel classification process in the complete absence of HC effects; this kind of personnel classification is referred to as allocation.

a. When the AVs are used as the basis of an optimal assignment process and the AVs are standardized so as to have the same mean and standard deviation, the HC effect is absent.

b. The allocation process capitalizes on the aptitude differences within an individual. The HC process capitalizes on the aptitude differences across individuals and the differences among jobs in terms of the value of receiving higher scoring individuals.

c. HC effects that can be capitalized on by either hierarchical layering or a LP process to increase *MPP* are present whenever jobs vary: (1) in validity or value, or (2) in variance or range of criterion scores expressed in terms of performance or value. When job performance is differentially weighted to express the value of jobs, the AVs must be comparably weighted (from an independent data source) if the HC process is to reflect unbiased job values.

3. While the presence of HC effects can provide a gain in MPP over chance assignment in the absence of allocation effects, and allocation effects can provide a gain in MPP in the absence of HC, the presence of both allocation and HC often provides very little increase over the presence of only allocation or only HC (Johnson & Zeidner, 1991, pp.29-41).

a. Very little loss in CE occurs when PP variables used as AVs (having the variance of the squared validities) are converted to standard score form with equal means and variances across all AVs; thus, although aptitude area scores used by the service lack HC effects, only a small gain in *MPP* occurs by converting the AAs to PPs.

b. It appears that HC and allocation effects are usually competitive in nature; capitalization on one largely prevents capitalization on the other.

4. A *g* type measure is *not* the most effective single measure to use in a pure HC assignment process; a measure based on the first H_d (PC type) factor is the best, in terms of classification efficiency, by a considerable margin (Statman, 1993).

a. The principal component (PC) solution obtained in the JP-C space sequentially maximizes H_a, component by component; the first component can be considered an approximation of g but with the additional advantage that it provides a higher average squared validity in the "back" sample JP-C space than is obtainable in any other way.

b. A PC solution obtained in the JP-C space and then orthogonally rotated to transform this solution into one which sequentially, factor by factor, maximizes H_d, is referred to as the H_d factor solution.

c. In the first model sampling experiment that compared use of the first PC with the first H_d factor in a simulation of a HC process, the *MPP* provided by the H_d factor scores weighted by validities was 1.7 times as large as that provided by PC scores weighted by validities (Statman, 1993).

5. A one-stage strategy in which selection and classification are accomplished simultaneously is theoretically superior to a two-stage strategy in which selection is accomplished in the first stage and classification is accomplished in the second stage. The one-stage strategy, unlike the two-stage strategy, assures that no one rejected for entry into the organization has a higher predicted performance for *any job* than *anyone* assigned to that job, and will theoretically yield a higher *MPP* for the combined selection and classification process. This theoretical superiority has been verified in a model sampling experiment by Whetzel (1991).

a. The traditional two-stage system uses a "g" type measure to effect selection in the first stage and tailored test composites to effect classification and assignment in the second stage.

b. The findings of Appendix C lead to the conclusion that the more applicants with lower g scores rejected in the first stage, the more potential there is for increasing *MPP* through optimal classification in the second stage.

c. The one-stage strategy calls for using tailored test composites to simultaneously reject some applicants and to classify and assign those who are selected; this is best accomplished by using the dual LP algorithm referred to as multidimensional screening (MDS). MDS uses, for each job, separate cut scores that will either reject a prescribed percentage of the population or maintain a specified quality level in terms of predicted performance.

d. MDS, a dual LP algorithm, calls for finding and implementing separate cut scores for the tailored test composites corresponding to each job family. A primal-dual LP computer program provided by ARI--in which a quota can be assigned to the rejection category without stipulating each individuals PP score for this "family"--will provide the same simultaneous selection and assignment batch solution as does the MDS algorithm. This primal-dual algorithm cannot directly provide a line-by-line optimal assignment solution but can provide the dual constants required by the MDS algorithm.

e. Once dual constants have been obtained it is possible to consider applicants one-at-a-time and still make optimal assignments. This is what the Air Force calls sequential assignment. *MPP* attributable to the separate effects of selection and classification is readily computed (analytically) when a two-stage strategy is utilized.

f. The separate contributions of selection and classification cannot be identified when MDS is utilized in a one-stage system.

6. A two-echelon strategy provides an effective basis for compromising between the desire to design a S/C system which will maximize *MPP* in optimal assignment and the desire to have a smaller number of AVs with intuitive content significance for career counselors and soldiers (Zeidner & Johnson, 1991, pp. 10, 160-162, 179-187, 205; Statman, 1993).

a. The first echelon of either a one-stage system or the second stage of a two-stage system provides for maximizing the MPP resulting from optimal initial assignment. To this end, there should be as many job families and corresponding FLS composites serving as AVs as can be supported by the validity data. The weights defining these AVs are entered into a computer-based system or "black box" and this first echelon assignment process is essentially invisible to both recruits and assignment personnel.

b. The second echelon uses a smaller number of tailored test composites, small enough in number to be easily recorded on a soldier's form 20 (and possibly discussed by counselors). Factor scores that are LSEs of factors rotated to simple structure in JP-C space can provide content-meaningful AVs. The meaningfulness of these factor scores can be determined by extending the rotated factors initially defined in the JP-C space into the predictor space using a process equivalent to the Dwyer factor extension method (1937).

7. Selection efficiency is proportional to validity when the applicant streams for each job are mutually independent.

a. Under this assumption the selection of one applicant has no effect on the number of applicants available for other jobs.

b. If, and only if this assumption is true, tailored tests are superior to a measure of *g* only to the extent that the validity of the tailored tests are higher than that provided by *g*.

8. Selection efficiency (SE), for a system in which selection for multiple jobs is accomplished from a common stream of applicants cannot validly be measured in terms of predictive validity; measurement of SE for such a system (one based on a common applicant source) must be accomplished through the use of a method which would also be appropriate for use in measuring classification efficiency (CE).

a. If cut scores are separately designated for each job or job family as the function of validity with respect to the job criterion, applicants are rank ordered on a measure of predicted validity (PP). Selection for the jobs having the highest cut scores is accomplished first. The process can be conceptually equivalent to hierarchical layering and the value of *MPP* can closely approximate what would be obtained from the use of a one-stage LP selection-classification procedure (MDS), if, and only if, the cut scores are equal to those which correspond to the dual parameters used in the MDS algorithm. Alternatively, if the cut scores are as low as are typical in the Army's personnel system, the *MPP* results will be much lower than could be obtained by an LP algorithm and even lower than obtainable by hierarchical layering.

(1). If the PP scores are a function of a single test (e.g., a measure of *g*) and the validity coefficients of *g* against each of the separate jobs or job families, *MPP* can be computed analytically.

(2). If the PP scores are a function of the tailored test composites and of the validities associated with each job or job family, a simulation of the selection system is required to obtain *MPP*, since the intercorrelations of the PP variables will affect the results.

b. Assuming the same cut scores as in (8a) above, but using a procedure in which applicants are considered for jobs either in a randomly selected order or by using a procedure in which selected individuals are assigned randomly to jobs, the hierarchical classification effect is eliminated. The increase in *MPP* provided by tailored test composites over that provided by the use of *g* is due to a more favorable selection ratio (SR) resulting from use of tailored test composites that have a population intercorrelation of less than 1.0.

E. Criterion-Related Principles

1. Criterion vs. Predicted Criterion Scores. Criterion and FLS scores clearly have the same correlations with any member of the set of predictor variables used to compute the LSEs of the actual criterion scores (e.g., the FLS composites used in the evaluation process). Thus predicted performance (PP) scores, defined as FLS composites, are substitutable for performance scores in computing the validities of assignment variables (AVs). The substitutability of predicted criterion scores for actual criterion scores is not applicable only to performance scores. It also applies to measures of retention rate, reenlistment rate, disciplinary offenses, etc., used as criterion scores. Substitutability extends to the computation of classification efficiency, as measured by mean predicted criterion scores, when making optimal assignment to multiple jobs.

a. The value of *MPP* computed from Brogden's formula as a function of LSE type AVs (R), the intercorrelations of theses AVs (r), and the number of jobs to which optimal assignment is directed, is the same whether computed using validities based on the actual criterion or based on predicted criterion scores.

b. Brogden (1946, 1951) contended that: (1) LSEs of the criterion variables are the most effective AVs for use in optimal assignment algorithms when the goal is to maximize *MPP*; and, (2) the average criterion score (e.g., *MPP*) resulting from the use of such AVs in optimal assignment models will be equal to the mean of either the criterion or predicted criterion scores. Subsequently, Brogden provided a rigorous proof of his earlier contentions regarding the substitutability of criterion and predicted criterion scores (Brogden, 1955). Over a decade later, Abbe (Dec., 1968) conducted a model sampling experiment that demonstrated the robustness of Brogden's proof with respect to the latter's assumptions. Her model sampling results were "consistent with Brogden's theoretical proof of equality of the two measures for infinite samples." (p 11)

c. Abbe (Dec., 1968) also showed that the use of LSEs instead of the actual criterion scores, as AVs in an optimal assignment (i.e., LP) algorithm, does not introduce bias. She tested the null hypothesis that mean differences between performance scores and LSEs of performance scores are the same before and after optimal allocation of personnel to jobs. This null hypothesis was accepted for samples ranging from n = 100 to l = 1,000, while maintaining an equality among the sums of the sample.

d. The substitutability of predicted criterion scores for actual criterion scores is essential to the credibility of DAT research paradigms. This principal assumption of DAT already has theoretical credibility

2. Criterion Components. Five criterion components were selected for use in the Army's Project A concurrent research effort. Only one of these, "core technical" proficiency for specific jobs, demonstrated its usefulness as the criterion variable in classification research (Wise, McHenry & Campbell, 1990). They state that, "For generalization across jobs, within each criterion factor, one equation fit the data for four of the five performance components. Different prediction equations were required for the component that reflects proficiency on the technical tasks specific to each job." p.355

a. When the same criterion components are available for each job, each job for which adequate data exist can be (at least in theory) uniquely represented by "best" weighted criterion components. The component weights of this composite might be obtained as the result of expert judgement (Sadacca, Campbell, DiFazio, Schultz, & White, 1990). Alternatively they could be obtained by using expert judgements to obtain scores representing the worth to the Army of different levels of performance on each job and then obtaining LSEs of job worth in terms of the criterion components. LSEs of predicted job worth could then be obtained as test scores in a test battery. Job worth from one source could be used as the selection and assignment variables, while an LSE derived from an independent source could be used as the evaluation variable.

b. The job specific, job worth composite consisting of "best" weighted criterion components would maximize, for a linear model, the prediction of job worth. When this best weighted composite becomes the dependent variable, the tests replace the criterion components as the independent variables, This LSE becomes the predicted criterion variable whose properties were described in DAT principle E1, its usefulness as a critical variable in a classification model then becomes clear.

c. Either job worth, predicted in terms of the tests, or predicted performance can be divided into two orthogonal components, both in joint space. The LSE resides in the intersection overlap of predictor and criterion variables expressed as vectors in the union of the predictor and criterion Euclidian spaces--one component which makes no contribution to CE, and one component which maximizes CE when utilized as the figure of merit for selecting tests and computing regression weights for AVs. The first of these orthogonal components is in every way equivalent to Brogden's g. That the Brogden g content of AVs makes no contribution to CE was proven by Brogden (1964); Zeidner and Johnson (1994) showed that the lack of contribution of g to CE can also be proven using the relationships in Brogden's 1959 model. This issue is discussed further in Appendix A.

d. There are <u>no</u> predictors or criterion variables relevant to classification that are not also relevant to selection; there are <u>both</u> predictors and criterion variables that are so highly loaded with Brogden's *g* that they are relevant to selection but not to classification (see Appendix A in Johnson, Zeidner, & Scholarios, in press).

3. Effect of Non-linear Relationships Between Criterion Variables and Job Worth. An LSE of job worth, in terms of predictor tests, can provide a measure of potential benefit to the organization that can be used as a figure of merit for evaluating both selection and assignment systems. The use of a constant in an LSE of job worth can raise or lower the mean predicted benefit score to correspond to the variation of job worth across jobs. However, non-linearity of relationships between LSEs of performance criteria and the true worth of each level of criterion values intuitively appears likely with respect to some jobs. Research on the linearity issue is clearly warranted. The state-of-the art for accomplishing an unbiased investigation of this issue needs improvement; practical methods for operationally implementing non-linearity considerations are even more inadequate at this time.

The major state-of-the art issues are on the criterion side. The greatest need is for a measure of benefit to the organization that reflects the worth of each level of performance (job worth) and that can adequately serve as the dependent variable for LSEs in which the independent variables are tests in an operational test battery. Adequacy of the measure of job worth should be determined in terms of its appropriateness for modifying the performance measure serving as the benefits measure in a utility model. A suitable measure of job worth must reflect both current and future policy and have reliability across expert judgements of job worth. Other constraints on how measures of the benefits to the organization (based on both the incumbents' performance and job worth) are obtained and used are discussed below.

a. The relationship between performance criteria or predicted criterion variables and job worth (for a specified criterion variable) can be logically placed in the following four categories. Categories 1 and 3 are appropriate for use in selection research while categories 2 and 3 are appropriate for use in classification research. Note that classification research also requires the presence of differential validity across assignment targets (see Appendix A in Johnson, Zeidner & Scholarios, in press).

(1) Jobs having a critical criterion score with little advantage accruing, in terms of job worth, for having employees who exceed this score.

(2) Jobs providing little differentiation regarding job worth of employees having criterion scores below a critical value, but providing considerable differentiation (i.e., linear prediction of benefits to the organization that reflect job worth) above this critical score.

(3) Jobs with a linear prediction of job worth over the entire criterion range.

(4) Jobs for which job worth is very poorly predicted by the specified criterion, that is, the criterion variable has a trivially low relationship to job worth.

b. A considerable amount of literature on this topic supports the general applicability of the linearity assumption for category (3). Where either category (1) or (2) appear credible with respect to specific jobs and criterion variables, the category (4) alternative must also be seriously considered (and evaluated).

c. The LSEs of job worth based on one or more predicted criterion composites can be expressed as weighted test composites. Because the test composite predicting job worth yield approximately normal scores in the population, additional credence can be given to the prediction that job worth is also approximately normally distributed is also for almost all jobs. This initial bell shaped distribution of predicted job worth scores can be readily modified to reflect the distribution shapes implied in (1) and (2).

d. Job worth is almost always determined judgmentally by policy makers; a change in the occupancy of policy making positions can radically change policy bearing on the determination of job worth. It is highly likely that the implementation process for an operational system utilizing job worth scores would span the sequential tours of two or more policy makers. Thus an improvement in the state-of-the art on the design of operational systems incorporating job worth desperately needs further research on two major issues: (1) the probability that job worth policy would remain stable across tours; and, (2) the reliability of judgements required to implement policy.

e. The use of the same source for job worth data to provide both improved AVs that reflect job worth and estimates of benefit scores based on LSEs of job worth can obviously inflate the resulting estimate of utility because of correlated error. Correlated error could be avoided only if the measures of job worth were completely error free. Note that direct measures of benefits reflecting job worth can be used only as the dependent variables for computing LSEs of benefits, not as AVs. LSEs of either performance or benefits can be utilized as both AVs and evaluation variables in multi-job situations, provided that the proper independence of data sources is maintained. DAT research paradigms which maintain independence among analysis, evaluation, and cross samples provide protection from all types of correlated error we have identified in the research studies conducted by the authors. However, we have not yet conducted a research effort involving the use of job worth as the measure of utility.

f. A two stage model, one in which the initial stage is selection into the organization, and the subsequent stage is one in which new recruits are classified and assigned to jobs, permits the use of specialized criterion variables separately for selection and classification research, and for the separate evaluation of each stage. The optimal first stage criterion is one which has maximum prediction of either performance or benefits (taking job worth into consideration) determined collectively against selective jobs representative of the average of every job in the organization. The optimal second stage (classification) criterion variable can then be one which has the effects of Brogden's *g* minimized.

g. The "best" figure of merit for use in conjunction with a one stage selection/classification process (in which selection and classification are simultaneously optimized) has not been fully determined as of this date. However, it is clear that the optimal figure of merit, for use in developing and evaluating system characteristics to maximize *MPP* obtained by optimal simultaneous selection and assignment, must contain an appropriate mix of the separate evaluation variables described in the above paragraph. Because, as far as we know, no operational one-stage optimal system has been installed anywhere in the world, the resolution of this issue is not an urgent one.

4. Alternative Shapes of Criterion Distributions

a. The scores of composite criterion variables, i.e., weighted composites of components, converge in statistical theory towards normality, even when some of the criterion components have bi-nomial or other non-normal distributions; predicted criterion scores, where the independent variables are tests with approximately normally distributed scores, converge even more rapidly toward normal distribution.

b. DAT based model sampling techniques for generating criterion scores constrained to have projected validity coefficients for each predictor variable necessarily have normal distributions because of the "central limits theorem" (Parzen, 1960; Fisz, 1963). Conversion to scores which have prescribed non-normal distributions, but which retain the desired validity coefficients, requires methods unique to the shape of each criterion distribution.

c. Some criterion distributions, as when the figure of merit is expressed as a rank order or as a percentile, have rectangular distributions. The generation of such criterion scores for use in a model sampling experiment requires that the designated population validity vector be multiplied by the appropriate value prior to the initial generation of normally distributed scores. A method for obtaining the appropriate multiplier has been described by Boldt (1965).

d. Some criterion variables can be scored only dichotomously--yes/no, present/absent, 1/0, etc. For such variables the figure of merit may be expressed in terms of the probability that each individual or entity will receive one of these scores (as in retention studies). The generation of synthetic scores taking on values of 0 or 1 is accomplished by using a normal curve table to convert into scores the probability that each entity will be associated with one dichotomous direction vs the other.

5. Quality Distribution. Over time, the Army has had a number of different policies stipulating varying levels of personnel quality as goals when acquiring personnel for job families or jobs. The Army has always defined quality in terms of such test scores as AFQT or its predecessor, AGCT. Other measures of quality proposed or considered at various times include

aptitude area scores, predicted performance in specific jobs or job families, general mental ability (*g*), and ability to read Army training manuals.

Responsibility for meeting desired personnel quality distribution goals across jobs rests initially with recruiters. The next opportunity to meet personnel quality goals comes during the classification and assignment process. Research investigators interested in the design of S/C systems must concern themselves with the effect that manipulation of S/C variables has on the meeting of personnel quality distribution goals and with the costs that may accrue from the use of the alternative strategies and systems. Fortunately, placing quality distribution constraints on assignment strategies and algorithms does not pose as great a problem on the use of optimal assignment algorithms as it might first appear (Nord & Schmitz, 1991).

a. The primal and dual linear programming (LP) algorithms, with special attention given to the dual parameters, provide two useful models from which a number of relationships among S/C system variables can be surmised. It is the weighted sum of all job AV means which comprises the objective function for the primal solution. The effect of S/C system variables on the obtained value of the objective function of an LP primal algorithm, used to accomplish optimal assignment of personnel, can be approximated in this manner:

Maximizing the overall mean AV score results in a wide dispersal of AV means across individual jobs. Relationships among such system variables as job quotas, intercorrelations among AVs, validities of AA composites for jobs, and cut scores, can be determined from examining a number of implications (discussed below) presented by the dual version of the LP algorithm (Johnson & Zeidner, 1991, pp. 48-50):

(1) Other things equal, the more unequal the total sets of quotas, the smaller the value of the overall objective function; and the smaller the quota for a particular job, the larger the mean AV score for that job. The expected mean AV, and, by implication, the personnel quality achievable for a large job family, could be increased in an optimal assignment model by shredding the family into two or

more job families. However, this shredding would be expected to have various side effects on other S/C variables impacting on job quality.

(2) Because the operational AA composites have a particularly high average correlation with AFQT, one should expect to find that the combat arms MOS, which are contained in two comparatively large job families, would have below-average AFQT scores (i.e., lower personnel quality) after optimal assignment.

(3) Again, other things equal, the higher the intercorrelations among jobs in a job family, the higher the sum of the mean AV scores (one for each job) for that job family. Intuitively, the decrease in the average intercorrelations among two AVs decreases this competition among the corresponding job families for the available quality.

(4) When the AVs are predicted performance variables, the higher the validity of the AV, the higher will be the mean AV score for a given job. If quality were also measured in terms of AV scores, higher validity for an AV would lead to higher quality personnel who, on the average, will have been assigned to jobs within the job family corresponding to the AV--usually the more technical jobs--on which performance is more predictable.

(5) The jobs with more valid AVs are also very likely to have higher quality-goals. The jobs with higher minimum (cut) scores will also have higher job quality-goals for a given job, regardless of how quality is measured (AFQT, *g*, AA, PP, etc.). Also, the validity of the AVs, when they are predicted performance variables, show high relationships among operational cut scores, quality goals assigned by policy makers, and the quality scores obtained from optimum assignments (Nord & Schmitz, 1991).

(6) The high correlation across jobs with high quality goals and the obtained mean AVs of jobs greatly reduces the need for remedial action to obtain the desired pattern of job quality. Most methods for meeting quality goals would greatly reduce the overall *MPP* obtainable from optimal assignment. However,

shredding of large families for use in an optimal assignment system, especially if the intercorrelations among AVs could be reduced at the same time, provides a method that could improve the meeting of quality goals while increasing overall *MPP*.

b. After selection and optimal assignment takes place, using AVs with equal expected means and standard deviations (as with the Army's operational AA composites), the hierarchical layering effects will have been eliminated and quality means will not be effected by the differing validities of the AVs. However, when measured in terms of predicted performance (even when personnel are assigned randomly after initial selection), the quality found in each job will be to some degree proportional to the validity of each job's performance predictor used in the evaluation process.

Conclusions

This sourcebook is a consolidation of concepts (including theoretical and methodological discussions) and results (from both model sampling experiments and analytical models) bearing on research or operational issues relating to selection and classification systems. Only studies in progress or completed prior to 1994 were used to develop principles. The Sourcebook details all DAT principles defined to date.

The Sourcebook permits the examination of selected principles that can be evaluated in the context of all other DAT principles and that focuses on broad DAT-based theory rather than on comparatively narrow operational issues.

An early version of DAT principles appeared in a chapter describing DAT (Johnson & Zeidner, 1991, pp.208-211). Only ten principles were defined in that chapter and the amount of evidence based on analytical models supporting the principles far outweighed the amount of evidence that was based on the use of model sampling results. Knowledge of DAT principles has at least tripled during the past three years, and a similar rate of growth is anticipated in the near future.

References

Abbe, E. N. (1968). Statistical properties of allocation averages. Memorandum 68-13. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A079 341)

Boldt, Robert F. (1965). A technique for simulating mixed normal and uniform distributions. Research Memorandum 65-9. Washington, DC: U.S. Army Personnel Research Office.

Brogden, H. E. (1946). An approach to the problem of differential predictions. Psychometrika, 11, 139-154.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. Educational and Psychological Measurement, 11, 173-196.

Brogden, H. E. (1954). A simple proof of a personnel classification theorem. Psychometrika, 19, 205-208.

Brogden, H. E. (1955). Least squares estimates and optimal classification. Psychometrika, 20, 249-252.

Brogden, H. E. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. Educational and Psychological Measurement, 19, 181-190.

Brogden, H. E. (1964). Simplified regression patterns for classification. Psychometrika, 29, 181-190.

Brogden, H. E., & Taylor, E. K. (1950). The dollar criterion--applying the cost accounting concept to criterion construction. Personnel Psychology, 3, 133-154.

Cronbach, L. J. (1979). The Armed Services Vocational Aptitude Battery--A test battery in transition. Personnel and Guidance Journal, 57, 232-237.

Fisz, M. (1963). Probability theory and mathematical statistics. New York: Wiley.

Fredericksen, N. (1968). Toward a broader conception of human intelligence. American Psychologist, 41, 445-452.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Gulliksen, H. (1987). Theory of mental tests. Hillsdale, NJ: Erlbaum.

Harris, R. N. (1966). A model sampling experiment to evaluate two methods of test selection. Research Memorandum 67-2. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A079 331)

Horst, P. (1954). A technique for the development of differential prediction battery. Psychological Monographs, 68(9, whole No. 380).

Horst, P. (1955). A technique for the development of a multiple absolute prediction battery. Psychological Monographs, 69(9, Whole No. 380) 1-22.

Horst, P. (1956). Multiple classification by the method of least squares. Journal of Clinical Psychology, 12, 3-16.

Humphreys, L. G. (1962). The organization of human abilities. American Psychologist, 17, 475-483.

Humphreys, L. G. (1979). The construct of general intelligence. Intelligence, 3, 105-120.

Hunter, J. E. (1983). The prediction of job performance in the military using ability composites: The dominance of general cognitive ability over specific aptitudes. Rockville, MD: Research Applications.

Hunter, J. E. (1984). The prediction of job performance in the civilian sector using the ASVAB. Rockville, MD: Research Applications.

Hunter, J. E. (1985). Differential validity across jobs in the military. Rockville, MD: Research Applications.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E., Crosson, J. J., & Friedman, D. H. (1985). The validity of the Armed Services Vocational Aptitude Battery for civilian and military job performance. Rockville, MD: Research Applications.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. Journal of Social and Biological Structures, 7, 93-118.

Jensen, A. R. (1986). *g*: Artifact or reality? Journal of Vocational Behavior, 29, 301-331.

Jensen, A. R. (1991). Spearman's *g* and the problem of educational equality. Oxford Review of Education, 17, 169-187.

Johnson, C. D., & Sorenson, R. C. (1974). Model sampling experimentation for manpower planning. In D. J. Clough, C. G. Lewis, & A. L. Oliver (Eds.), Manpower planning models (pp. 37-52). London: English Universities Press.

Johnson, C. D., & Zeidner, J. (1991). The economic benefits of predicting job performance. Vol. 2: Classification efficiency. New York: Praeger.

Johnson, C. D., Zeidner, J., & Leaman, J. A. (1992). Improving classification efficiency by restructuring Army job families (Tech. Rep. 947). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A250 139)

Johnson, C. D., Zeidner, J., & Scholarios, D. (1990). Improving the classification efficiency of the Armed Services Vocational Aptitude Battery through the use of alternative test selection indices (IDA Paper P-2427). Alexandria, VA: Institute for Defense Analyses.

Johnson, C. D., Zeidner, J., & Scholarios, D. (in preparation). Developing new test election and weight stabilization techniques for designing classification efficient composites. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Leaman, J. A. (1992, August). Restructuring job families to improve potential classification efficiency. Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

Mosier, C. I. (1951). Problems and designs of cross validation. Educational and Psychological Measurement, 11, 5-11.

Nord, R., & Schmitz, E. (1991). Estimating performance and utility effects of alternative selection and classification policies. In J. Zeidner, & C. D. Johnson (Eds.). The economic benefits of predicting job performance. Vol. 3: The gains of alternative policies. New York: Praeger.

Olson, P. T., Sorenson, R. C., Hayman, K. W., Witt, J. M., & Abbe, E. N. (1969). Summary of SIMPO-I model development (Research Rep. 1157). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD 692 790)

Parzen, E. (1960). Modern probability theory and its applications. New York: Wiley.

Ree, M. J., & Earles, J. A. (1991). The stability of g across different methods of estimation. Intelligence, 15, 271-278.

Ree, M. J., & Earles, J. A. (1994). The ubiquitous predictiveness of *g*. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), Personnel selection and classification: New directions (pp. 127-135). Hillsdale, NJ: Erlbaum.

Sadacca, R., Campbell, J. P., Difazio, A. S., Schultz, S. R., & White, L. A. (1990). Scaling performance utility to enhance selection/classification decisions. Personnel Psychology, 43, 367-378.

Schmidt, F. L., Hunter, J. E., & Larson, M. (1988). General cognitive ability vs. general and specific aptitudes in the prediction of training performance. Some preliminary findings (Contract No. Delivery Order 0053). San Diego, CA: U.S. Navy Personnel Research and Development Center.

Scholarios, D. M. (1992, August). A comparison of predictor methods for maximizing potential classification efficiency. Paper presented at the American Psychological Association Annual Meeting, Washington DC.

Scholarios, D. M., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. Journal of Applied Psychology, 79, 412-424.

Sorenson,, R. C. (1965). Optimal allocation of enlisted men--Full regression equations versus aptitude area scores (Tech. Research Note 163). Washington, DC: U.S. Army Personnel Research Office.

Spearman, C. (1904). General intelligence objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spearman, C. (1927). The abilities of man. New York: Macmillan.

Statman, M. A. (1992, August). Developing optimal predictor equations for differential job assignment and vocational counseling. Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

Statman, M. A. (1993). Improving the effectiveness of employment testing through classification: Alternative methods of developing test composites for optimal job assignment and vocational counseling. Unpublished doctoral dissertation, The George Washington University, Washington, DC.

Thorndike, R. L. (1985). The central role of general ability in prediction. Multivariate Behavioral Research, 20, 241-254.

Thurstone, L. L. (1935). The vectors of mind. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Welsh, J. R., Jr., Kucinkas, S. K., & Curran, L. T. (1990). Armed Services Vocational Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Air Force Systems Command.

Whetzel, D. L. (1992, August). Multidimensional screening: Comparison of a single-stage personnel selection/classification process with alternative strategies. Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 3, 23-40.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. Personnel Psychology, 43, 335-366.

Zeidner, J. (1987). The validity of selection and classifiction procedures for predicting job performance (IDA Paper P-1977). Alexandria, VA: Institute for Defense Analyses.

Zeidner, J., & Johnson, C. D. (1991a). The economic benefits of predicting job performance, Vol. 1: Selection utility. New York: Praeger.

Zeidner, J., & Johnson, C. D. (1991b). The economic benefits of predicting job performance, Vol. 3: Estimating the gains of alternative policies. New York: Praeger.

Zeidner, J., & Johnson, C. D. (1992). Classification efficiency and systems design. Journal of the Washington Academy of Sciences, 81, 110-128.

Zeidner, J., & Johnson, D. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), Personnel selection and classification: New directions (pp. 377-410). Hillsdale, NJ: Erlbaum.

Appendix A

The Effect on PCE of the Removal of Brogden's *g* from AVs

(In the Context of Personnel Classification)

The effect on potential classification efficiency (PCE) of removing *g* from a predictor set by the deletion of tests from LSE composites can be inferred, apart from the use of model sampling techniques, by several different analytical methods. This appendix will emphasize the different effects reduction in Brogden's *g* is expected to have on PCE, but the effect reducing other kinds of *g* can be expected to have on PCE will also be discussed.

There are at least four approaches to measuring a general aptitude component sometimes referred to as *g*. While it is often claimed that it makes very little difference as to which of these measures is used, we know that at least one of these four types of *g*, Brogden's *g*, is very different from the others. Because it is Brogden's *g* that plays a central role in Brogden's MPP model (1959) and is the *g* construct that is most clearly extraneous (or possibly harmful) to the classification process when included in AVs, this is the *g* which should most interest researchers interested in personnel classification.

Spearman's 2-factor structure is implied by Brogden's assumptions for his *MPP* model. Brogden's *MPP* model requires a general factor defined in the joint predictor-criterion space which has equal validities for all jobs. While we do not find much use for his specific factors, except for use in the derivation of his model, a factor with the prescribed properties of his *g* has considerable usefulness and can be credibly identified in most data.

We see no harm in considering most factorial measures of g that are obtained in total test or common factor test space as falling into a common category. We will essentially ignore such versions of psychometric *g* for the purposes of this appendix. There are two other types of g, in addition to Brogden's *g*, that are instead determined in the joint predictor-criterion space. One of these is the principal axis solution of the covariances among predicted performance measures, and the other is a classification efficient solution (differential validity factor) that is only an orthogonal rotation away from the principal axis factor solution in joint space. The differential validity factor method provides a solution that successively maximizes Horst's squared differential validity. Both of these latter measures of g are described and studied in Statman (1993, pp 120-123) and the latter is appropriately referred to by Statman as the "differential validity factor".

Brogden's g can be computed from the covariances of the predicted performance measures, **C**, by finding the largest equal elements of a column vector, **g**, such that a residual matrix equal to **C** - **g g'** is positive semidefinite. If there are sufficient degrees of freedom, **g** can be directly computed without using an iterative method. (Johnson and Zeidner, 1991, pp 159-162). Statman (1993, pp 286-288) found in her study that the elements in **g** (all elements are equal) was .317, compared to a range of .415 to .876 for the elements of the largest principal axis factor and a range of -.330 to .378 for the classification efficient differential validity factor. Statman used the same **C** matrix to obtain all three factor estimates of g.

The CE obtainable from using only the largest factor of a principal axis solution of the covariances of predicted performance measures as the AV is investigated by Statman (1994). She compares the CE of this measure of g with the classification efficient single factor of the PP covariances. The diagonal terms of this covariance matrix contain the squared multiple correlations with the criterion variable that is associated with each AV.

Brogden's *MPP* model (1959) estimates classification efficiency as a function of $f(m)\ R\ (1-r)^{1/2}$. We refer to $R\ (1-r)^{1/2}$ as our estimated classification efficiency (ECE). Constructing examples within the constraints of Brogden's model in which AV validities for the corresponding job criterion variable all have the same value, R, and the intercorrelations among all AVs are equal to r, we see that reducing the factor loading on Brogden's g has no effect on the value of ECE. A reduction in g decreases both R and r, each by amounts that leaves ECE invariant.

Continuing to use ECE as our measure of PCE, but leaving the factor structure unspecified except for the presence of a Brogden g, we obtain a different estimate of the effect of Brogden's g on ECE. The removal of Brogden's g from the joint predictor-criterion space decreases both R and r by the squared factor loading on **g** ,or g^2, Using an example in which the factor loading on **g** is .35, R = .40, and r = .85, we obtain a value for ECE of .3873. Removing the effect of g yields: R = .2775, r =.7275, and ECE = .5220. We see that, for this example, there is a 35 percent increase in ECE that resulted from the deletion of Brogden's *g* from the space spanned by the predictor set.

Brogden (1964), without making use of the assumptions of his 1959 model, proves that when the AVs include all the tests in the battery and are best weighted, the adding of the same constant to the weights of each test will not change the *MPP* obtained through optimal assignment. If the addition of a selected constant to a given test for all AVs results in a zero multiplier for that test across all AVs, the adding of such a constant is clearly equivalent to deleting the test from the predictor set. For such a deletion to occur, the deleted test must closely approximate Brogden's *g*.

Most Army empirical data yields considerable hierarchical classification (HC) efficiency from the use of psychometric *g* to optimize assignments to multiple jobs. It may surprise some investigators that when a single factor score from each individual is used to make these assignments, Statman's model sampling results (1993, p. 157) show that use of the classification efficient factor provides a 171 percent superiority in *MPP* (.144 vs. .084) over that provided by the principal axis solution in joint space. Whetzel (1991, pp. 90-91) found that psychometric *g* provided only 11 percent of the CE (measured in terms of *MPP*), of that provided by using tailored test composites as the AVs in an optimal assignment process--after all *MPP* values were corrected for the effect of the selection process. When HC is removed from her measure of psychometric *g* in joint space, the amount of *MPP* after correcting for *MPP* due to the selection process is only .014, close enough to zero to confirm the hypothesis that psychometric *g* standardized to eliminate HC effects, like Brogden's *g*, makes little or no contribution to classification efficiency.

Appendix B

The Effect on the Standard Errors of Regression Weights of Removing *g* From AVs

The standard error of estimate for a regression weight is commonly written as follows:
$s_e = \{[1 - (R_{z\bullet 1,2,3....n})^2] / [1 - (R_{1\bullet 2,3....n})^2]\}\ f(N,m).$
$f(N,m) = [1/(N-m)].$

We will consider only the ratio to the left of f(N,m) in considering the reduction of *g* by the selective removal of tests from a set of predictors that are then best weighted to form an AV composite. The predictor variables designated in this formula by the numbers from 1 to n must be the tests that are empirically constructed rather than artificial constructs formed from the basic tests, such as either oblique or orthogonal factors. If this representation of s_e held for regression weights applied to factor scores to predict performance, the denominator of the ratio would be increased to 1.0 for every factor in a composite of weighted factors, thus appearing to minimizing s_e. However, the derivation of this formula does not justify its use with respect to variables that are not directly measured. In this appendix we are only considering regression weights that have been applied to independent variables consisting of the tests that were actually administered to the examinees.

We would reach very different conclusions as to the effect of *g* on s_e depending on whether we are referring to psychometric or Brogden's *g*. The important differences between these two measures of *g* are discussed in Appendix A.

It may be useful to describe the relationships among the independent and dependent variables in terms of a specified factor structure. We will separately consider s_e in the context of three factor models that provide very different intuitive results. However, some intuitive observations apply to all three factor structures. For example: (1) in any practical situation, the removal of *g* from the predictor set will necessarily reduce the magnitudes of both multicorrelation coefficients in the above ratio; (2) *g* will be expected to correlate higher, on the average, with the tests in the predictor set than will any test (unless that test is a pure measure of g); (3) g will correlate higher with a performance criterion than will any test (again, unless that test is a pure measure of g).

If the predictor variables were structured in accordance with Spearman's 2-factor model, the denominator of the above ratio would, for all tests in the predictor set, converge to one as g is completely removed from the set of predictors. The numerator would decrease at a lesser rate and does not converge on 1.0 because the loadings on the specific factors explain an appreciable amount of the prediction of the criterion. To the extent that the loadings of tests on g are negatively correlated with the loadings on the specific factors, the numerator of the ratio would not be greatly affected by the removal of g from the predictor set.

If only one factor is required to explain the reliable variance in the joint predictor-criterion place, the regression weight is equal to zero after the removal of g, and there is no point to computing the s_e after such an event. However, when there is unidimensionality, the only relevant dependant variable (the z in our formula) for a regression model is g itself. It is clear that s_e will be lowest in such a model for predictors having the purest loadings of g. A pure loading in g for a test requires that all of the reliable variance of that test be perfectly correlated with g. If unidimensionality exists in the test space, as well as in the joint space, the reliability, validity, and intercorrelations--thus also the s_e--would all be a function of the loading of g on the predictors.

Because non-zero allocation efficiency cannot exist in the context of the unidimensional model, we are tempted to favor Spearman's 2-factor model over a unidimensional model, in applying intuition, as to whether: (1) relying on g, or (2) minimizing g, is the better strategy for forming a predictor set--when the goal is to reduce the magnitude of s_e. However, we prefer the use of a group factor model in the joint space as our preferred factor model for this purpose. Preliminary analyses using this model indicate that when two alternate variables differ primarily in their loadings on g, the predictor variable with the smaller loading on g can be expected to have the less stable weight.

Appendix C

The Effect of Different Selection Ratios on *MPP*

(In the Context of Personnel Classification)

This appendix explores the relationship between SR and CE. Using both model sampling results, and an analytical model with input drawn from credible "made-up" examples, we will provide evidence that the effect varying SRs has on the magnitudes of validities and the intercorrelations of the assignment variables (AVs) should logically have the opposite effect on an estimate of CE than the same changes in SR would be expected to have on SE in a second stage.

Compare a two stage selection procedure with a two stage selection/classification model in which the first stage is selection into the organization and the second stage provides classification and assignment of personnel to multiple jobs. It is clear that the greater the number rejected in the first stage, the lower the validity of predictor variables (R) in the second stage--assuming the predictors are at least moderately correlated with the variable used for the initial, first stage, selection process. Selection efficiency (SE) in a second stage will be decreased by lowering the SR (i.e., increasing the rejection rate), but the effect that varying the SR has on classification efficiency (CE) is much more complicated.

Based on results from a model sampling experiment, the effect of the size of SRs on CE (e.g., *MPP*s after optimal assignment to jobs) appears to be either zero or in the opposite direction for a second classification stage. This is seen to contrast with the general effect of SR on SE in the first stage. (Whetzel, 1991, pp 90-91). An improved understanding of these somewhat surprising results can be had from noting which of our "made up" examples, processed through our analytical model, provide results that are comparable to her model sampling results. This understanding comes from noting that our model is very sensitive to certain characteristics in our examples.

Whetzel (1991) conducted a model sampling experiment in which selection into the organization was simulated by deleting all artificial individuals in each input sample that have a selection variable *score* (i.e., for psychometric *g*) below a specified score that has the expectation of yielding a specified SR. Considering only the *MPP* values remaining after subtracting out the amount of *MPP* that would have resulted from selection alone, in order to make comparisons across two different values of SR, Whetzel retained only the *MPP* that could be attributed to CE. She conducted model sampling experiments in which two strategies are separately simulated: (1) a one stage (simultaneous selection and classification) and (2) a two stage (selection and then classification) selection and optimal assignment system. Results for two major facets were provided: (1) SR = .75 vs. SR = .50; and, (2) one stage vs. two stage strategies.

Within the two-stage strategy, an SR of .5, as compared to an SR of .75, had no more than a trivial superiority in CE after deletion of the *MPP* amount explainable as due to selection. However, there was a non-trivial increase in *MPP* for the condition in which SR was equal to .5, as compared to .75, when using the one stage strategy (and deleting the amount of *MPP* explainable as due to selection). The difference in *MPP* across the two values of SR was .029. Whetzel (1991, pp. 90-91)

In our use of an analytical model to demonstrate the effect of SR on classification efficiency, we will use restriction in range formulae to compute, for three examples, the effect on validities and intercorrelations of a truncation of the distribution of selection variable scores at a cut score corresponding to an SR value of .75 and .5. Note that this procedure involves correcting from a unrestricted population to a restricted group, in contrast to the more common correction that corrects a correlation coefficient computed on a restricted sample in order to estimate coefficients that would be obtained in an unrestricted group.

We use Brogden's *MPP* function, $MPP = f(m)\, R\, (1\text{-}r)^{1/2}$, as an estimate of classification efficiency in the "back" sample. We ignore f(m), a function of the number of jobs to which optimal assignment is targeted, reducing his estimate of CE to a function of the validity of assignment variables, i.e., R) and the intercorrelation among the assignment variables (i.e., r). We will refer to this estimate of CE as ECE for a column heading found in Table 2. We will compare values of ECE obtained using restricted and unrestricted correlation coefficients.

Formula 19 (Gullikson, 1987, p. 149) is used to compute both R and r corrected for the specified truncation effects of a specified SR. We use our own notation in formula 19 and simplify to take advantage of our special assumptions. The following notation pertains to our procedure:

r_{xa} = unrestricted correlation coefficient between the selection variable (x) and an assignment variable (a). Note that, in our model, all AVs have the same correlation with x, and with the criterion variable, y.

In our model we identify two AVs, a and b, thus,

$r_{ax} = r_{bx}$, and $r_{ay} = r_{by}$.

r_{xy} = unrestricted correlation coefficient between the selection variable and the criterion variable corresponding to each job. In our model all criterion variables are assumed to have the same validity coefficient.

r_{ay} = unrestricted correlation coefficient between each AV and the corresponding job criterion variable. As with r_{xy}, all criterion variables are assumed to have the same validity coefficient for all AVs. Also, as noted above, $r_{ay} = r_{by}$.

Using the above notation and modifying the input, we write formula 19 to separately define R and r as follows:

$r = [r_{ab} + J\,(r_{xa})^2] \;/\; [1 + J\,(r_{xa})^2\,]$

$R = r_{ay} + J\,r_{xy}\,r_{xa} \,/\, G;$

$G = \{\,[\,1 + J\,(r_{xa})^2\,][\,1 + J\,(r_{xy})^2\,]\,\}^{1/2}$

$J =$ $[(s_x \,/\, S_x)^2 - 1]$, where S_x is the unrestricted SD of the selection variable represented in statistical standard score form (S_x = 1.0), and s_x is the SD of the selection variable after truncation on the lower tail of x as represented by SR = .75 or SR = .5. The formula for computing s_x is derived, presented, and discussed in Appendix D.

For SR = .75, $(s_x)^2$ = .534737, and for SR = .5, $(s_x)^2$ = .363515. The parameters of the three examples are provided in Table 1, and the *R*, *r*, and ECE values for both SR conditions for each of these three examples are provided in Table 2.

Table 1
Three Examples used as Model Input

Examples	r_{xy}	r_{ay}	r_{ab}	r_{ax}	ECE
1	.30	.40	.80	.90	.1789
2	.30	.35	.75	.85	.1750
3	.35	.40	.75	.80	.2000

Note. ECE = estimate of classification efficiency

Table 2
Results Obtained after Selection with Indicated SR

Examples	SR	*R*	*r*	ECE
1	.50	.338	.587	.217
	.75	.355	.679	.201
2	.50	.263	..537	.179
	.75	.290	.623	.178
3	.50	.269	.578	.175
	.75	.644	.331	.198

Note. ECE = estimate of classification efficiency

In defining each of our three examples, we stipulated that: $r_{ax} > r_{ab}$ and $r_{ay} > r_{xy}$. Assuming that *x* is a good measure of *g*, each of the AVs should correlate higher with *x* than with the other AVs. Also, each of the AVs should have at least some superiority over *g* with respect to its ability to predict the criterion variable corresponding to that particular AV. These two hierarchical relationships are clearest in example 1, have been decreased in example 2, and further decreased in example 3.

We find for example 1 that ECE is largest for SR =.5, next largest for SR = .75, and smallest for SR = 1.0. For example 2, the values of ECE are almost equal across the three SR conditions. For example 3, SR equal to 1.0 provides a trivial superiority in ECE as compared to SR equal to .75, but a definite superiority when compared to SR equal to .5. Although not shown here, we found that the improbable examples in which r_{xy} is equal, or superior, to r_{ay} provide ECE values under the condition in which SR is equal to 1.0 are far superior to conditions in which SR is set equal to .75, which in turn are far superior to conditions in which SR is equal to .5.

Appendix D

Computing Techniques Required in Determining Selection Ratio Effects

The squared standard deviation of a distribution of scores (e.g., x) can be usefully expressed as the mean squared x score minus the squared mean of x. We will first consider a group of scores for which the mean is zero and the standard deviation is equal to one. If we delete all scores below the mean and compute a new SD on the remaining scores, the squared mean of these remaining scores remains equal to one, regardless of the shape of the total distribution of scores. Thus the standard deviation of the remaining scores is equal to one minus the squared mean of the remaining scores. If the scores in our example are normally distributed, the squared SD for a SR of .5 is equal to $1 - (z/p)^2$, where p is equal to the SR (.5) and z is the ordinate of the normal curve at the mean ($x = 0$). When x is not equal to zero, a more complex formula for computing the squared SD on the (remaining scores after a truncation of the normal curve at the x value associated with the specified SR) is required. A more generic formula for computing the squared SD for all scores above any value of $k = x$, when all scores below k are deleted (rejected), is derived below.

In our derivation, we first express the mean square of x, defined as f below, in terms of a normal distribution:

$$f = [\mathrm{D} \int_k^{\infty} x^2 e^{(-x^2/2)} \, dx \,] \, / \, p.$$

Note that D represents a constant that can be ignored in the integration process. Accomplish the required integration of the normal curve function, f, by writing f in terms of $u \, dv$, $u = x$ and $v = e^{(-x^2/2)}$. Making use of the traditional integration by parts formula, $\int u \, dv = uv - \int v \, du$, we obtain the following result: $f = 1.0 + k \, (z/p)$, where z is the ordinate of the normal curve at the point where the abscissa, x, is equal to k, and p is equal to the area under the normal curve to the left of k. The desired variance of a distribution of standard scores, distributed like the normal curve and truncated at the abscissa score of k, is equal to $f - (z/p)^2$. Referring to a table for the normal curve we find that for SR = .5, k is equal to zero, p is equal to .5, z is equal to .3989, and $(s_x)^2$ is equal to .363515; for SR = .75, p = .75, k is equal to minus .674516, z is equal to .31775, and $(s_x)^2$ is equal to .534737.